LWG
LOWER WILLAMETTE GROUP

PORTLAND HARBOR SUPERFUND SITE ECOLOGICAL RISK ASSESSMENT:

INTERPRETIVE REPORT: ESTIMATING RISKS TO BENTHIC ORGANISMS USING PREDICTIVE MODELS BASED ON SEDIMENT TOXICITY TESTS

DRAFT

MARCH 17, 2006

DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

Prepared for:
The Lower Willamette Group

Prepared by:

Windward Environmental LLC
Avocet Consulting
TerraStat Consulting Group

WE-06-0002

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AET	apparent effects threshold
ANOVA	analysis of variance
AOPC	area of potential concern
ASTM	American Society for Testing and Materials
AWQC	ambient water quality criteria
COPC	chemical of potential concern
CSL	cleanup screening level
DDT	dichloro-diphenyl-trichloroethane
Ecology	Washington State Department of Ecology
EPA	US Environmental Protection Agency
ERA	ecological risk assessment
FPM	floating percentile model
GIS	geographic information system
ISA	initial study area
LAET	lowest apparent effects threshold
2LAET	second-lowest apparent effects threshold
LEL	lowest effect level
LOE	line of evidence
LRM	logistic regression model
LWR	Lower Willamette River
MDD	minimum detectable difference
NOAA	National Oceanic and Atmospheric Administration
ODEQ	Oregon Department of Environmental Quality
PAH	polycyclic aromatic hydrocarbon
PCA	principal components analysis
PCB	polychlorinated biphenyl
PEC	probable effects concentration
PEL	probable effects level
PEL-Q	probable effects level quotient
Programmatic Work Plan	Portland Harbor Remedial Investigation/Feasibility Study Programmatic Work Plan
QA	quality assurance
QA/QC	quality assurance/quality control
RI/FS	remedial investigation/feasibility study
RL	reporting limit
RM	river mile

ROD	Record of Decision
SEL	severe effects level
SPI	sediment profile imagery
SQG-Q	sediment quality guideline quotient
SQS	sediment quality standard
SQV	sediment quality value
TEC	threshold effects concentration
TEL	threshold effects level
TOC	total organic carbon
TRV	toxicity reference value
TZW	transition zone water
USACE	US Army Corps of Engineers

EXECUTIVE SUMMARY

This report presents the results of an evaluation of existing sediment quality values (SQVs) and site-specific predictive models that could be used in assessing risk to benthic invertebrates in the ecological risk assessment (ERA) for the Portland Harbor Superfund Site.

The overall objective of the study was to develop a predictive toxicity model that would characterize the relationship between sediment chemistry and benthic invertebrate toxicity. The recommended model will be used to identify the primary chemicals of potential concern (COPCs) that may cause toxicity to benthic invertebrates and provide site-specific SQVs. These SQVs will be used to predict potential toxicity to benthic invertebrates and identify areas that may pose unacceptable risk to benthic communities.

The reliability of five existing sets of SQVs was assessed, but none were found to have acceptable reliability in predicting benthic toxicity in Portland Harbor. Consequently, further exploratory analyses were conducted, and site-specific models were developed. Two principal models, the floating percentile model (FPM) and the logistic regression model (LRM), were chosen to determine if a predictive relationship between sediment chemistry and benthic invertebrate toxicity response could be developed for Portland Harbor. In addition to these two models, site-specific apparent effects thresholds (AETs) were developed and evaluated for use as potential SQVs.

The FPM showed that the relationship between chemicals and toxicity varied by effects endpoint. The *Hyalella* mortality and *Chironomus* growth and mortality endpoints were sensitive to similar chemicals and had strong relationships with bulk hydrocarbons, polycyclic aromatic hydrocarbons (PAHs), ammonia and sulfides, certain metals (e.g., cadmium, mercury, silver), and certain other organics (hexachlorocyclohexane, polychlorinated biphenyls (PCBs), dichloro-diphenyl-trichloroethane (DDTs), chlordane, and di-n-butyl phthalate). The *Hyalella* growth endpoint had strong relationships only with percent fines and ammonia and weak relationships with a few metals such as copper, arsenic, nickel, and zinc. Based on the FPM approach, specific areas with potential benthic toxicity that are related to known upland sites and sources within Portland Harbor were identified along both banks of the river. The results of this approach correspond well both with measured toxicity and with the conceptual site model.

The LRM also showed that chemicals associated with toxicity vary by endpoint. For the *Chironomus* pooled (mortality and growth) endpoint, the strongest relationships exist with diesel-range hydrocarbons, PAH-like compounds (i.e., carbazole and dibenzofuran), sulfide, certain metals (i.e., lead and mercury), and certain other organics (DDE, chlordane, and di-n-butyl phthalate). For the *Hyalella* mortality endpoint, the strongest relationships exist with diesel-range hydrocarbons and residual-range hydrocarbons (i.e., bulk hydrocarbons), PAHs (e.g., naphthalene), sulfide, and certain

other organics (hexachlorocyclohexane, chlordane, DDE, and total DDTs). The *Hyalella* pooled endpoint had the strongest relationships between toxicity and percent fines, ammonia, sulfide and certain metals (i.e., aluminum, selenium, copper, and mercury). The LRM also identified specific areas with potential benthic toxicity similar to those identified by the FPM model.

The FPM is recommended for assessing risk to benthic invertebrates in Portland Harbor. This approach will provide the most comprehensive set of site-specific SQVs to identify areas of potential benthic toxicity within the harbor. Although initially, both the FPM and LRM showed promise in predicting Portland Harbor-specific toxicity based on surface sediment concentrations, the analysis indicates that the FPM would better meet the needs of the RI/FS being conducted for the Portland Harbor Superfund Site. The error rates for this model were lower than those for the LRM, and the FPM provides a more complete set of site-specific SQVs. Although the LRM is not recommended, the results from this model confirm the areas of potential benthic toxicity identified by the FPM. The site-specific AETs were found to have low reliability in terms of a high false negative rate and are not proposed for use.

1.0 INTRODUCTION

The assessment of risk to benthic organisms is an integral part of the ecological risk assessment (ERA) approach as outlined in the *Portland Harbor Remedial Investigation/Feasibility Study Programmatic Work Plan* (Programmatic Work Plan) (Integral et al. 2004). This report presents the results of sediment toxicity testing and the derivation of sediment quality values (SQVs) for the Portland Harbor Superfund Site, hereafter referred to as the Study Area. These elements form the primary line of evidence (LOE) to be used in assessing risk to benthic invertebrates in the ERA within the Study Area.

Several LOEs were identified in Appendix B of the Programmatic Work Plan to provide empirical information for estimating risks to benthic invertebrate communities. The primary LOE addresses benthic toxicity either by laboratory exposure of benthic organisms to contaminated sediment or predicted toxicity based on the observed relationship between laboratory toxicity and sediment chemistry. Supporting LOEs were also identified in the Programmatic Work Plan to provide additional information for discussing the results of the primary LOE. The supporting LOEs include the comparison of benthic invertebrate tissue residue concentrations to toxicity reference values (TRVs) and comparison of surface water and transition zone water (TZW) chemical concentrations to ambient water quality criteria or other appropriate screening values. These supporting LOEs will identify which pathways may contribute risk to benthic populations and to the benthic community in general.

The direct measure of toxicity is based on standard laboratory toxicity tests using *Chironomus tentans* and *Hyalella azteca*, which measure the effects of sediments on growth and mortality of the test organisms. Both of these species, or closely related species, are indigenous to the Lower Willamette River (LWR). They will also serve as surrogates in the baseline ERA for the natural invertebrate community in the river because of their abundance and distribution throughout the river and their importance as prey to many other species. This is a standard and widely accepted approach for evaluating adverse effects from contaminated sediments.

The LOEs serve two broad purposes in the remedial investigation/feasibility study (RI/FS) process for the Study Area. The primary purpose will be to estimate site-wide risks to the benthic community. The toxicity data will be used as the primary LOE. Where there is no toxicity data, site-specific SQVs will be used to predict risk to benthic communities. When the predictive model is used, stations with unacceptable risk to benthic invertebrates will be identified as those with sediment concentrations of chemicals of potential concern (COPCs) greater than their respective site-specific SQVs.

The second use of the LOEs will be to support the identification of areas of potential concern (AOPCs) in the ERA, which will be used to delineate SMAs, which will in turn be evaluated for remedial action in the feasibility study. AOPC will be delineated based

on sampling locations where any of the bioassay endpoints exceed their toxicity thresholds or contaminant concentrations exceed site-specific SQVs (actual bioassay results will take precedence over SQVs), as well as other risk results. Tissue, surface water, and TZW samples will be used as supporting LOEs for confirming potential risk areas and evaluating transport pathways.

The predictive relationships presented in this report have been developed for key chemical contaminants found in Portland Harbor sediments and for benthic effects endpoints that are relevant to the risk decisions for the Study Area. The resulting SQVs will provide EPA and its partners with a way to assess the potential for risk to the benthic community associated with direct toxicity and to define sediment that may be affected. The direct toxicity data and the SQVs will be important in determining the need for implementing sediment remedial actions.

1.1 STUDY OBJECTIVE

As stated in Windward (Windward 2005a), the overall objective of the analysis presented in this report is to develop a predictive toxicity model that characterizes the relationship between sediment chemistry and benthic invertebrate toxicity in the LWR. The resulting model will be used:

- To derive SQVs that are sufficiently reliable for predicting benthic toxicity within the Study Area
- As one LOE for identifying areas where chemical concentrations in sediment may pose a risk to benthic invertebrates

Predictive models are intended to identify threshold concentrations of contaminants associated with toxicity to benthic invertebrates based on demonstrated relationships between direct measures of toxicity using the standard toxicity tests and surface sediment chemistry. As stated earlier, the resulting SQVs will be used to evaluate the potential for toxicity in locations where direct measures of toxicity are not available. This general approach has been or is being adopted by other jurisdictions (e.g., the states of Washington, Florida, and California) to develop criteria for managing potential risks from contaminated sediment. Site-specific predictive toxicity models have been previously used by EPA at other Superfund sites (e.g., Calcasieu Estuary, Louisiana, and Commencement Bay, Washington).

1.2 OVERVIEW OF STUDY APPROACH

This section describes the analytical approach taken to evaluate the predictive nature of the relationship between the sediment chemistry and the benthic toxicity data collected in the Study Area.

- The first step was to ensure the quality of both the chemistry and toxicity data sets and to organize and prepare both data sets for analysis (as detailed in Section 2.0). Following validation (Integral 2005b; Windward 2005b), the chemical data set was organized to be more useful for the exploratory modeling process. For example, chemicals with fewer than 30 detected values were excluded from further analysis because this appeared to be the minimum threshold for a usable distribution for the development of SQVs based on analyses of other data sets from Oregon and Washington (Avocet 2003). Polycyclic aromatic hydrocarbons (PAHs), dichloro-diphenyl-trichloroethanes (DDTs), polychlorinated biphenyls (PCBs), chlordane isomers, endosulfan isomers, and dioxin-like compounds (i.e., PCBs, furans, and dioxins) were summed to represent totals. After validation of the toxicity data set, biological effects levels for each toxicity test endpoint were selected, and the value of incorporating regional toxicity data sets was explored.
- An analysis of the ability of five existing SQV sets (CCME 2002) to predict toxicity in the Study Area was initiated (see Section 3.0). Although these existing SQVs incorporate data from the Pacific Northwest for similar species and habitats, they did not reliably predict toxicity in the Study Area. The details of this analysis are provided in Appendix A.
- An exploratory analysis evaluating the relationships among chemical concentrations as well as between chemical concentrations (including percent fines and total organic carbon [TOC]) and biological effects levels (both magnitude of bioassay response and response [hit/no-hit] classifications) was conducted to determine the efficacy of developing site-specific SQVs (see Section 4.0).
- Site-specific data were then applied to two candidate models to develop SQVs based on the relationship between sediment chemistry and benthic invertebrate toxicity (as detailed in Section 5.0). The two models, the floating percentile model (FPM) and the logistic regression model (LRM), were identified in the *Portland Harbor RI/FS Ecological Risk Assessment: Estimating Risks to Benthic Organisms Using Sediment Toxicity Tests* (Windward 2005a). Model development involved numerous iterations to reflect varying levels of biological effects and different ways to incorporate the four individual endpoints (i.e., retaining individual test results or pooling growth and mortality for both *Chironomus* and *Hyalella*), based on peer review and agency input. The reliability of the models was evaluated using several reliability parameters. In addition to these two models, site-specific SQVs were developed for those chemicals not included in the FPM using the apparent effects thresholds (AETs) (PSEP 1988), and the reliability of these AETs was evaluated.

- The major findings of the study are summarized in Section 6.0. The findings include methods, endpoints, and effects levels that were evaluated for use but not proposed as part of the final model, the strengths and weaknesses of the two principal models (i.e., the FPM and LRM), and the proposal of a site-specific set of SQVs.
- Recommendations are provided with respect to which model should be selected to identify areas where chemical concentrations in sediment could pose risks to benthic organisms within the Study Area (see Section 7.0).

1.3 DOCUMENT ORGANIZATION

This report follows the approach presented in the *Portland Harbor RI/FS. Ecological Risk Assessment: Estimating Risks to Benthic Organisms Using Sediment Toxicity Tests* (Windward 2005a), which was submitted to the US Environmental Protection Agency (EPA) in January 2005. The remaining sections of this document present the exploratory analysis, including model development, to be used for assessing risk to benthic invertebrates, as follows:

- Section 2.0 –Details the assessment of data quality and organization of both the sediment chemistry data and the toxicity data.
- Section 3.0 – Evaluates the reliability of existing SQVs.
- Section 4.0 – Discusses the exploratory analyses performed to understand the relationship between sediment chemistry and toxicity data.
- Section 5.0 – Presents the development of candidate benthic toxicity prediction models and site-specific AETs.
- Section 6.0 – Summarizes the results of the modeling efforts.
- Section 7.0 – Presents recommendations for model selection and use.
- Section 8.0 – Lists cited references.

Supporting information is presented in Appendices A through E.

2.0 DATA QUALITY AND ORGANIZATION

This section presents an overview of data quality and organization and data reduction rules for both toxicity and chemistry data. Surface sediment samples were collected in the LWR from July 19 through November 5, 2004 (Round 2), at a total of 521 stations (Integral 2005a) (see Figure 2-1). The majority of the stations (515) were within the Study Area (River Mile [RM] 2 to RM 11). The remaining six stations were located upstream of the Study Area between RM 16 and RM 25. Chemical analyses were performed on all surface sediment samples from the 521 stations for all analytes except butyltins, petroleum, and dioxins and furans. Butyltin data were available for 110 stations, petroleum data for 203 stations, and dioxin and furan data for 104 stations. Toxicity testing using *Chironomus tentans* and *Hyalella azteca* was performed on 233 surface sediment samples, including the six ambient upstream stations. The predictive models were developed based on the 233 samples with co-located sediment chemistry and bioassay data. Based on the models, the risk to benthic communities was predicted at the 282 stations with chemistry data only. In addition, chemical analyses of other surface sediment samples collected in Round 1 and 2 were included in the risk assessment, bringing the number of sediment samples with chemistry data to 396.

Data organization and reduction steps were performed on both the sediment chemistry and toxicity data to allow for more efficient exploratory data analyses and predictive toxicity model development. Steps were taken to remove chemicals with limited detections, sum some chemical groups into a single value (e.g., total PCBs), and prepare the toxicity data by determining hit/no-hit designation for each endpoint and each effects level (see Section 2.1.2). The toxicity data set is discussed in Section 2.1; the chemistry data set is discussed in Section 2.2.

2.1 TOXICITY DATA

This section presents the evaluation of the toxicity data, including QA/QC, options for combining Portland Harbor Round 2 data with historical data, and biological effects definitions.

2.1.1 Quality Assurance

The toxicity data underwent an extensive QA/QC process, including validation by a third party. Data were deemed to be of excellent quality and fully usable for any future application (Windward 2005b).

2.1.2 Biological Effects Definitions

Before modeling the relationship between sediment chemistry and benthic invertebrate toxicity, the different levels of biological response needed to be defined. The biological effects levels used in the analyses are intended to correspond conceptually to "no effects level" (Level 1), "minor effects level" (Level 2), and "moderate effects level" (Level 3). As requested by EPA (EPA 2005a), the three levels were set at 90, 80, and 70% of the

response observed in the control sediment, respectively. Use of these three levels divides the overall data set into four categories according to the severity of effects. Table 2-1 presents the definitions of the three effects levels, and Figures 2-2 and 2-3 present the "hit/no-hit" designations for the 233 stations for each of the two toxicity tests (*Chironomus* and *Hyalella*, respectively) and the three effects levels.

Table 2-1. Definitions of biological effects levels

Test and Endpoint	Hit/No-Hit Criteria for Effects Levels[a]		
	Level 1 (90%)	Level 2 (80%)	Level 3 (70%)
Hyalella azteca 28-day mortality	T/C < 0.9	T/C < 0.8	T/C < 0.7
Hyalella azteca 28-day growth	(C - T)/C > 0.1	(C - T)/C > 0.2	(C - T)/C > 0.3
Chironomus tentans 10-day mortality	T/C < 0.9	T/C < 0.8	T/C < 0.7
Chironomus tentans 10-day growth	(C - T)/C > 0.1	(C - T)/C > 0.2	(C - T)/C > 0.3

a To be considered a toxic sediment at each of the three levels, the test response must also be statistically different from the negative control response ($p < 0.05$).

T – mean of untransformed mortality or weights in test sediment

C – mean of untransformed mortality or weights in negative control sediment

The biological effects levels are based on statistically significant differences from the negative control in addition to minimum difference thresholds (Table 2-1). The decision to use the negative control in the comparison was made in cooperation with EPA and its partners because of the greater reliability observed using this approach, the fact that standardized freshwater reference sites are not yet available in the region, and because the results are more conservative (Ecology 2002).[1] At any of these effects levels, a toxicity test endpoint response is considered a hit if the difference in response is greater than the defined threshold and is statistically different from control; a no-hit station has a difference less than the threshold or is not statistically different from control. If the observed difference exceeds the threshold but is not statistically significant, the test must have had a minimum detectable difference (MDD) equal to or less than the threshold. Indeterminate stations were defined as those that had actual differences that exceeded the threshold, non-significant statistical results, and an MDD greater than the threshold. MDDs were determined for each sample comparison using post-hoc power analysis with 80% power, one-tailed $\alpha = 0.05$, and the sample variances. This process ensured that large-magnitude differences were not designated as no-hits based on lack of statistical significance due to low power. Figures 2-2 and 2-3 present the locations of indeterminate stations for *Chironomus* and *Hyalella,* respectively.

The no effects level (Level 1) was initially defined based solely on a statistically significant difference from negative control. However, evaluation of the statistical power of the significance tests indicated that many samples would be labeled

[1] Reliability: correct predictions/total stations.

"indeterminate," thereby removing them from further analyses. In consultation with EPA and its partners, it was determined that a minimum threshold difference between site stations and negative control was needed to identify non-toxicity beyond a statistical difference.

Therefore, the no effects level was re-defined to require a minimum difference of 90% relative to control for both survival and growth. This definition ensured that very small magnitude differences were not defined as hits based solely on significance tests with very high statistical power.

The minor effects level (Level 2) and moderate effects level (Level 3) were based on an approach suggested by EPA, NOAA, and the Oregon Department of Environmental Quality (ODEQ 1999) for this project.

2.1.3 Use of Historical Toxicity Data

The initial analyses were performed on Portland Harbor Round 2A sediment chemistry and toxicity data. As an exploratory analysis to determine whether including historical data would improve model results, existing bioassay data from Portland Harbor were also added. However, the only available data that had passed QA requirements were from the *Chironomus* 10-day test, and the addition of these data did not measurably improve the model reliability. Some new data have recently been collected that would provide additional *Chironomus* and *Hyalella* data, such as the US Army Corps of Engineers (USACE) channel deepening data, but these data were not available in time for this report. Largely because of data completeness issues in the historical data sets and the lack of improvement in reliability that resulted from combining them with the Round 2 data, this report relies on the Round 2A data in developing a benthic model

2.2 CHEMISTRY DATA

The surface sediment chemistry data underwent an extensive quality assurance/quality control (QA/QC) process, including validation by a third party. The data were deemed to be of high quality (Integral 2005b).

2.2.1 Data Quality

A review of the synoptic sediment chemistry data collected in 2004 was performed to ensure that only data of acceptable quality were included in the exploratory analysis and model development. This review was based on the qualifiers assigned to each individual chemical concentration during the data validation process. All chemical qualifiers used in the Portland Harbor sediment chemistry data are presented in Table 2-2. Individual data points with the qualifiers presented in Table 2-3 were not included in the analyses.

Table 2-2. Qualifiers used in the Portland Harbor sediment chemistry data

QUALIFIER	DEFINITION
J	Estimate
JT	Combined qualifier
N	Presumptive evidence of analyte[a]
NJ	Combined qualifier (presumptive/estimate)
NJT	Combined qualifier (presumptive/estimate/average)
R	Rejected – failure to meet QA guidelines
T	Value is an average or selected result (following standard project rules)
U	Not detected at value shown
UJ	Combined qualifier (not detected/estimate)
UJT	Combined qualifier (not detected/estimate/average)
UT	Combined qualifier (not detected/average)

[a] Metals: the matrix spike sample recovery is not within control limits. Organics: tentative identification; the analyte exhibits low spectral match parameters but is present.

Table 2-3. Qualifiers that resulted in the exclusion of Portland Harbor sediment chemistry data

Qualifier	Definition
N	Presumptive evidence of analyte[a]
NJ	Combined qualifier
NJT	Combined qualifier
R	Rejected – failure to meet QA guidelines

[a] Metals: the matrix spike sample recovery is not within control limits. Organics: tentative identification; the analyte exhibits low spectral match parameters but is present.

Other data sets (from Rounds 1 and 2) included in the exploratory analysis underwent a similar QA evaluation. Each individual data point was evaluated based on the qualifiers assigned during the QA process by the original author(s), and results with qualifier definitions listed in Table 2-3 were excluded. The exclusion of data with the N-qualifier primarily affected the pesticide data. Between 23 and 53% of the data for the following pesticides were excluded: aldrin, hexachlorocyclohexane (alpha-, beta-, and delta-), nonachlor (cis- and trans-), dieldrin, and methoxychlor. Between 35 and 67% of the summed data of DDD, DDE, DDT, total DDT, total chlordane, and total endosulfan were excluded. In addition, 11% of the 1,2,3,7,8-pentachlorodibenzofuran data and 8% of the calculated total dioxin and furan data were excluded.

2.2.2 Data Organization and Reduction

All chemical data were used in the exploratory analysis of the relationship between sediment chemistry and toxicity data (Section 4.0). Specific exclusions of individual

chemical endpoints because of low detection frequencies, or exclusion of non-detected data, are noted for each of the analyses detailed in Section 4.0.

For the modeling efforts described in Section 5.0, only detected values were used because undetected chemistry values do not provide useful information for the development of a predictive relationship between sediment chemistry and benthic invertebrate toxicity. Chemicals with 30 or more detected values were included in the modeling efforts because this appeared to be the minimum threshold for a usable distribution for the development of SQVs based on analyses of other data sets from Oregon and Washington (Avocet 2003). SQVs were developed on a site-wide basis; however, chemicals with fewer than 30 detected values may be important when evaluating smaller areas or individual sources. Figures B-1 through B-4 (Appendix B) present the locations of chemicals with fewer than 30 detected values. Several of these chemicals cluster in areas that are related to known upland sites and sources along both banks of the river and will need to be considered when evaluating these specific areas.

After the exclusion of data with fewer than 30 detected values, the two modeling approaches had slightly different rules for including individual chemical endpoints. Within the FPM, the final SQVs are a function of the joint distribution of all chemicals present in the Study Area. The presence of non-toxic, naturally occurring crustal elements such as aluminum and selenium can confound the development of meaningful SQVs for the remainder of the analytes. Consequently, aluminum and selenium were excluded from the FPM. In the LRM approach, individual regression models are developed for each analyte independent of the concentrations of other analytes. In the final multi-chemical model, the contribution of non-toxic elements to the overall predictions of toxicity can be evaluated. Consequently, there is no harm in including highly correlated, non-toxic analytes in the LRM so selenium and aluminum were included.

Other analytes that are derived quantities (e.g., dioxin TEQs) and chemicals that are identified as highly correlated with each other (e.g., PAHs) are represented in the FPM as sums; they are included in the LRM as both individual chemicals and as sums. Certain conventional analytes, such as specific gravity and total solids, were screened out of both models because they are not considered contaminants. However, other conventional analytes, including percent fines, bulk sediment ammonia, and sulfides, were retained in the two models because of their apparently strong correlation with toxicity in some biological endpoints.

The data screened out due to the above factors are summarized in Table 2-4 and shown in Figures B-1 through B-4 (Appendix B). In some analyses, different rules were used to select analytes for inclusion; these analytes are presented where relevant in Sections 4.0 and 5.3.

Table 2-4. Analytes screened out prior to model development

Analytes with Fewer than 30 Detected Values[a]
1,2,3-Trichloropropane (1)
1,2,4-Trichlorobenzene (6)
1,2-Dichlorobenzene (5)
1,2-Dichloroethane (1)
1,4-Dichlorobenzene (5)
2,3,4,5-Tetrachlorophenol (5)
2,3,4,6/2,3,5,6-Tetrachlorophenol coelution (7)
2,4,6-Trichlorophenol (22)
2,4-D (6)
2,4-DB (1)
2,4-Dichlorophenol (2)
2,4-Dimethylphenol (1)
2-Chlorophenol (1)
2-Methylphenol (2)
4-Chloro-3-methylphenol (5)
4-Nitroaniline (1)
Acetone (4)
Aniline (8)
Benzene (19)
Benzyl alcohol (11)
Bis(2-chloroethyl)ether (2)
Carbon disulfide (12)
Chlorobenzene (13)
Chloroform (15)
Chromium, hexavalent (3)
Diethyl phthalate (7)
Dimethyl phthalate (12)
Di-n-octyl phthalate (5)
Endrin (11)
Endrin ketone (2)
Ethylbenzene (14)
gamma-Hexachlorocyclohexane (21)
Gasoline-range hydrocarbons (21)
Heptachlor (10)
Heptachlor epoxide (2)
Hexachlorobutadiene (21)
Hexachloroethane (26)
Isopropylbenzene (21)

Table 2-4. Analytes screened out prior to model development

m,p-Xylene (17)
MCPA (2)
MCPP (1)
MTBE (6)
Methyl ethyl ketone (20)
Mirex (4)
N-nitrosodiphenylamine (2)
o-Xylene (23)
Styrene (1)
Toluene (5)
Trichloroethene (6)
Crustal Elements and Analytes Not Related to Toxicity
Aluminum (in the FPM)
Selenium (in the FPM)
Specific gravity
Total organic carbon (covaried with percent fines)
Total solids
Correlated Individual Chemicals Replaced by a Sum[b]
Individual grain size parameters (replaced by percent fines)
Individual dioxins and furans (replaced by TEQ [see Section 5.3 for exceptions])
Individual DDD, DDE, and DDT isomers and sums (replaced by total DDTs)
Individual PAHs, LPAH, HPAH, dibenzofuran, and carbazole (replaced by total PAHs in most cases [see Sections 4.0 and 5.3 for exceptions])
Individual Aroclors (replaced by total PCBs)
Individual endosulfans (replaced by total endosulfans)
Individual chlordanes, nonachlors, oxychlordane (replaced by total chlordane [see Section 5.3 for exceptions])

[a] Analytes were detected at least once; the number of times detected is shown in parentheses. Any analyte not listed in the table and not retained for model development was never detected.

[b] Sums used for model development were consistent with sum definitions used throughout the Portland Harbor project.

2.2.3 Chemical Summation

For the model development, PAHs, DDTs, PCBs, chlordanes, endosulfans, and dioxin-like compounds were summed as totals according to the summation rules that have been established for the Portland Harbor RI/FS. These chemicals are often reported and evaluated as sums because they appear to express their toxicity more accurately on an additive basis. Using summations reduces covariance problems, and

past side-by-side comparisons of other Oregon and Washington data sets have shown better reliability when summations are used.

Total concentrations were calculated using the following summation rules:

- In samples where all chemicals contributing to the sum were detected, all detected concentrations were summed to represent the total concentration.
- In samples where some chemicals contributing to the sum were detected and some were not detected, only detected concentrations were summed to represent the total concentration.
- In samples where no chemicals contributing to the sum were detected, the highest detection limit was selected as the total concentration and was qualified as non-detected.

2.2.4 Normalization

Normalization of non-polar organic compounds and metals could be applied in an attempt to improve the reliability of the predictive model(s). However, no actual advantage has been revealed in past side-by-side comparisons of other Oregon and Washington data sets, and the reliability of the non-normalized sediment quality guidelines is generally the same or better than the normalized guidelines. Therefore, the data were not normalized to TOC or other variables.

3.0 COMPARISON TO EXISTING SEDIMENT QUALITY VALUES

The next step in the analysis was to evaluate if any existing SQVs in use in North America would be able to reliably predict toxicity to benthic invertebrates in the Study Area. If any of the existing SQV sets could be used to predict toxicity in Portland Harbor, it would not be necessary to develop site-specific models or SQVs. However, none of the existing SQVs were reliable for this purpose. A brief summary of the reliability analysis is provided below, and a complete discussion is presented in Appendix A.

- **TELs/PELs** – Threshold effects levels (TELs) are intended to represent chemical concentrations below which biological effects rarely occur. Probable effects levels (PELs) are intended to represent chemical concentrations above which adverse biological effects frequently occur. TEL/PEL values have been adopted in Canada and several states (CCME 2002).
- **TECs/PECs** – Consensus-based SQVs have been proposed by a group of private and agency sediment researchers in an attempt to unify the wide variety of SQVs available in the literature. They are similar in concept to TELs/PELs (Ingersoll et al. 2000; MacDonald et al. 2000). Threshold effects concentrations (TECs) and probable effects concentrations (PECs) have been used in Great Lakes areas of concern (MacDonald et al. 2000).
- **LELs/SELs** – The screening level concentration approach was developed by the Ontario Ministry of the Environment and is based on the presence or absence of benthic species in freshwater sediments (Persaud et al. 1993). The lowest effect level (LEL) corresponds to a level at which effects would be expected in only 5% of benthic species, while the severe effects level (SEL) represents a level at which effects would be expected in 95% of benthic species.
- **Washington Freshwater SQS/CSL** – The FPM was developed in an effort to improve the reliability of freshwater SQVs for Washington State (Avocet 2003; Avocet and SAIC 2002). Sediment quality standards (SQS) and cleanup screening levels (CSLs) have been calculated and are currently applicable to freshwater sediments in Washington State (Avocet 2003).
- **Quotient Methods** – Quotient methods were developed as an approach to increase the predictive ability of certain SQVs (Long et al. 1998) and have been applied to TELs/PELs and TECs/PECs (described above). Two quotient methods – one using sums and one using individual chemicals – were evaluated.

3.1 RELIABILITY ASSESSMENT METHODS

The seven reliability parameters are listed below and shown in Figure 3-1. These reliability parameters are used for the evaluation of existing SQVs, as well as for the reliability analysis of the two site-specific models and the site-specific AETs described in Section 5.0. Details of the chemical and biological data preparation methods are described in Appendix A.

- **False negatives** – Incorrectly predicted no-hits/total hits
- **False positives** – Incorrectly predicted hits/total no-hits
- **Sensitivity** – Correctly predicted hits/total hits
- **Efficiency** – Correctly predicted no-hits/total no-hits
- **Predicted hit reliability** – Correctly predicted hits/total predicted hits (this measure is equivalent to "1988 Efficiency" in Avocet (Avocet 2003; Avocet and SAIC 2002))
- **Predicted no-hit reliability** – Correctly predicted no-hits/total predicted no-hits
- **Overall reliability** – Correctly predicted stations/total stations

For each existing SQV set, the more protective of the two thresholds (i.e., TEL, TEC, LEL, and SQS) was compared to Levels 1 and 2, and the higher of the two thresholds (i.e., PEL, PEC, SEL, and CSL) was compared to Level 3, consistent with the narrative intent of these SQVs. Each of the four individual bioassay endpoints was assessed. In addition, a pooled endpoint was derived by combining all four endpoints from the two tests.

3.2 RELIABILITY ANALYSIS RESULTS

The reliability analysis results for the existing SQV sets are summarized in this section. Results for Levels 2 and 3 are presented in Tables 3-1 and 3-2, respectively; results for Level 1 are very similar to those of Level 2. A complete discussion of reliability results for the existing SQVs is presented in Appendix A.

Table 3-1. Reliability analysis for Level 2 biological effects using existing SQVs

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
***Chironomus* Growth**				
TEL	100	4	10	100
TEC	100	17	12	100
LEL	96	4	10	67
Washington SQS	83	46	14	96
***Chironomus* Mortality**				
TEL	100	2	15	100

Table 3-1. Reliability analysis for Level 2 biological effects using existing SQVs

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
TEC	97	14	16	97
LEL	97	1	14	67
Washington SQS	76	43	19	91
***Hyalella* Growth**				
TEL	99	4	42	67
TEC	92	19	44	72
LEL	100	5	42	100
Washington SQS	62	45	43	61
***Hyalella* Mortality**				
TEL	100	1	9	100
TEC	100	14	10	100
LEL	95	1	8	67
Washington SQS	80	42	12	96
Pooled Endpoint				
TEL	99	2	55	67
TEC	94	20	59	72
LEL	99	2	55	67
Washington SQS	66	49	61	54

LEL – lowest effect level
TEC – threshold effects concentration
TEL – threshold effects level
SQS – sediment quality standard
SQV – sediment quality value

Table 3-2. Reliability analysis for Level 3 biological effects using existing SQVs

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
***Chironomus* Growth**				
PEL	82	59	13	97
PEC	65	70	14	95
SEL	53	80	16	95
Washington CSL	65	54	9	95
***Chironomus* Mortality**				
PEL	68	57	16	94
PEC	56	68	17	93
SEL	52	79	23	93
Washington CSL	72	53	16	94
***Hyalella* Growth**				
PEL	44	56	19	80
PEC	31	66	17	79
SEL	31	80	25	82

Table 3-2. Reliability analysis for Level 3 biological effects using existing SQVs

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
Washington CSL	51	52	20	81
***Hyalella* Mortality**				
PEL	72	56	12	96
PEC	67	68	15	96
SEL	67	79	21	97
Washington CSL	83	53	13	97
Pooled Endpoint				
PEL	57	59	40	74
PEC	45	70	42	72
SEL	41	84	55	74
Washington CSL	61	55	40	74

CSL – cleanup screening level
PEC – probable effects concentration
PEL – probable effects level
SEL – severe effects level
SQV – sediment quality value

None of the existing SQV sets perform well enough to use them in predicting biological effects at the Portland Harbor Superfund Site. The lower thresholds (the TELs, TECs, and LELs) are far too conservative to be useful because they classify all or nearly all stations as hits (low efficiency). The higher thresholds (the PECs, PELs, and SELs) are more successful at predicting toxic effects, yet the error rates are still high enough that substantial portions of the Study Area could be incorrectly classified as contributing to adverse effects.

In general, the quotient methods are an improvement over most of the SQV sets discussed above although not sufficiently reliable for use in predicting toxicity results at this site (see Appendix A). It is possible that the quotient approach has merit, but it needs to be optimized on a site-specific basis. Overall, the reliability results for existing SQV sets and the quotient methods suggested that the development of a site-specific SQV set or predictive model was necessary to improve reliability and reduce error rates.

4.0 EXPLORATORY ANALYSES TO SUPPORT DEVELOPMENT OF SITE-SPECIFIC SQVS

Once it was determined that existing SQVs would not be good predictors of toxicity in Study Area sediments (see Appendix A), exploratory analyses of the potential relationships between sediment chemistry and toxicity were conducted to support the development of site-specific SQVs. Two types of exploratory analyses were conducted: simple statistical correlations among the chemistry and toxicity data and multivariate analyses. The resulting correlations among chemical endpoints were used to support the site-specific model development, such as providing supporting justification for the use of sums rather than individual chemical endpoints. The preliminary analyses also helped to provide an understanding of correlations among chemical and biological endpoints, since the site-specific SQV models are based on correlative and not causative relationships.

Multivariate analysis (i.e., cluster analysis and principal components analysis) helped identify geographic locations where groups of chemical endpoints were elevated and how those chemical concentration ranges related to observed toxicity in the area.

4.1 STATISTICAL CORRELATIONS

Pairwise scatter plots and correlation coefficients were used to illustrate and describe the relationships between chemical endpoints. The analysis used only data with 30 or more detected values and primarily focused on correlations within groups for metals, LPAHs, HPAHs, pesticides, and organotins and correlations between endpoints within other miscellaneous groups (e.g., phthalates, phenols, and miscellaneous organics). Because of skewness in the data set, the data were natural log-transformed prior to analysis. Most of the correlation coefficients were highly significant ($p < 0.01$) due to the large sample sizes and the strong trends. Figure 4-1 identifies the best correlative relationships between chemicals, defined as statistically significant Pearson's $r > 0.9$ (alpha = 0.01). Pearson's correlation coefficient was used to identify linear correlations that would justify the use of sums in place of individual analytes. Visual assessment of the pairwise relationships indicated that trends were linear when the skewed chemical endpoints were log-transformed.

Matrices of pairwise scatter plots (see Figure 4-2 for an example scatter plot; all scatter plots are presented in Appendix C) were developed for all data on the natural log scale. The plots in the first row and the plots in the first column are the same set of plots but the axes are switched. For example in Figure 4-2, antimony is on the x-axis of all plots in the first column and on the y-axis of all plots in the first row. The plots on the first row are, respectively: antimony (y) vs. arsenic (x), antimony (y) vs. cadmium (x), antimony (y) vs. chromium x), etc. Going down the first column, the plots are: arsenic (y) vs. antimony (x), cadmium (y) vs. antimony (z), etc. These plots were used to allow for quick review of all pairwise relationships among a group of chemicals. The

following general conclusions were made based on a review of the scatter plots (Figure 4-2 and Appendix C).

- Individual HPAHs are all highly correlated amongst themselves and naturally also with the HPAH sum. The sum could easily be used instead of individual HPAHs without much loss of information.
- Many of the LPAHs are highly correlated amongst themselves, and also with LPAH sum, although correlations are not as high as for the HPAHs (see Figure 4-2 and Appendix C). Napthalene and acenapthylene have slightly more variable patterns than some of the other individual LPAH analytes. Carbazole and dibenzofuran were included in with both the HPAH and LPAH plots because they had high correlations with several of the individual PAHs (Figure 4-1).
- Metals are not highly correlated amongst themselves, and there are some divergent patterns in the data (e.g., chromium vs. antimony).
- Individual butyltins are correlated: mono- and dibutyltin are highly correlated, and tetra- and tributyltin are highly correlated; although the two pairs are not well-correlated with each other. The set of detected data for tetrabutyltin is more limited than the other organotins (n = 34 vs. n = 70).
- There are limited correlations among pesticides; similarly, phenols are not well correlated amongst themselves and neither are phthalates.
- There are some correlation patterns between unrelated analytes from different groups (e.g., PCBs and cadmium, total chlordane and total DDTs). These are presented in Appendix C.

The relationship between chemistry (natural-log scale) and physical characteristics (percent fines and TOC; natural-log scale) were also investigated. The strongest correlations are presented in Figure 4-1, and examples of the kinds of relationships observed are presented in Figure 4-3. Percent fines was weakly correlated with most metals; however, correlations with the crustal metals such as aluminum and selenium were fairly strong ($r = 0.82$ and 0.70, respectively, on untransformed data). Even if correlations were not highly linear throughout the range, it was true for nearly all chemicals that high concentrations occurred in sediments with the highest fine-grained fractions (i.e., high concentrations implied high percent fines, but high percent fines did not always imply high concentrations).

Finally, the relationship between the magnitude of toxicity for the four individual endpoints and chemistry or physical characteristics (percent fines and TOC) was also investigated. The four toxicity endpoints were control-adjusted, and the chemistry data set was natural-log-transformed and included non-detects at the detection limit. This analysis was limited to only those chemicals that had greater than 50% detection frequency. The scatter plots did not reveal much correlation between the magnitude of

toxicity and these chemical and physical analytes, with a few exceptions, which are noted below and shown in Figure 4-3:

- *Hyalella* growth had a very slight negative correlation (i.e., lower growth with higher chemical concentrations) with many of the metals and with percent fines; the relationship was very flat (except for a few influential data points with no growth) for the non-metals.
- The three samples with no *Hyalella* growth (i.e., G288, G294-1, and G298, in which all test organisms died during the test exposure period) had the highest concentrations of the PAH sums, dibenzofuran, and diesel-range hydrocarbons.
- *Chironomus* endpoints had negative correlations with diesel-range hydrocarbons; *Hyalella* endpoints (both mortality and growth) displayed a weak relationship with diesel-range hydrocarbons.
- The growth endpoints for both *Hyalella* and *Chironomus* had very slight negative correlations with percent fines; the relationship for *Chironomus* growth was fairly variable.
- The ranges of responses for *Hyalella* mortality and the *Chironomus* endpoints were very narrow for most of the samples. There was no correlation between these biological endpoints and chemistry. The subset of samples that did have poor toxicity test responses (i.e., high mortality/low growth) for these endpoints did not have high chemical concentrations for any chemicals except diesel-range hydrocarbons.

4.2 MULTIVARIATE ANALYSES

Several multivariate approaches were used to assess and describe the relationships among chemical analytes and between chemical analytes and biological endpoints.

A cluster analysis was performed to group stations according to their chemical concentrations. The chemical variables used in the cluster analysis were based in part on an evaluation of the scatter plots (i.e., sums were used in place of individual congeners, and chemical analytes that correlated with toxicity were included) as well as a review of the detection frequencies. Missing values are not allowed in this analysis. In order to minimize the influence of the method for treating non-detects, the variables included in the final list were those that had at least 65% detection frequency, although the list also included hexachlorobenzene (53% detected) and selenium (48% detected). Non-detected values were included at the detection limit.

The chemical variables included in the cluster analysis were: aluminum, antimony, arsenic, cadmium, chromium, copper, lead, mercury, nickel, selenium, silver, zinc, bis-2-ethylhexyl phthalate, carbazole, dibenzofuran, hexachlorobenzene, total chlordane, total DDTs, total HPAHs, total LPAHs, total PCBs, and percent fines (TPHs were not included because the number of samples that had TPH results was fairly limited). Only

samples that had one or more detected concentrations for the specified analyte list were included, which resulted in a total of 231 samples for this analysis. Similarities were computed using Euclidean distance on the scaled data matrix (each value was standardized by subtracting the mean and dividing by the standard deviation for that variable; this puts all endpoints on the same scale). Several clustering algorithms were used to identify clusters; the final results shown used compact (furthest neighbor) linkage to attain clusters.

A principal components analysis (PCA) was done on the same chemical data set described above. The correlation matrix was used to place all endpoints on the same scale and reduce the influence of outliers.

A classification tree model was also used in an attempt to describe toxicity status based on chemistry. The response variable in these models was the hit classification described below, which is based on control-adjusted growth and survival. This exploratory modeling occurred prior to the establishment of the final effects levels used in the modeling process presented in Section 5.0, but they represent approximately the same levels of effects. The hit classifications were defined as:

- 0 for the best samples with > 90% control-adjusted survival and > 90% control-adjusted growth
- 1 for the intermediate samples with 75 to 90% control-adjusted survival and 70 to 90% control-adjusted growth
- 2 for the worst samples with < 75% control-adjusted survival or < 70% control-adjusted growth

The following conclusions were reached based on the multivariate analyses.

- Five components were required to explain at least 70% of the variability in the data set, indicating substantial heterogeneity in the combination of chemicals present.
- However, these differences in chemical constructs were along a continuum so that distinct clusters of stations did not exist.
- There was no relationship between the principal components based on chemistry and the biological endpoints.
- Sediment samples that appear to have fairly similar chemical constructs (e.g., a similar mixture of chemicals in similar concentration ranges) show a wide range of toxicity responses (i.e., some exhibit low toxicity, while others exhibit high toxicity).
- The classification tree model failed to identify consistent patterns in chemical concentrations for the toxicological responses.
- While grouping samples based on the sediment chemistry data resulted in a range of toxic responses, grouping the samples according to toxicity

revealed substantial heterogeneity in the chemistry associated with the toxic responses.

Appendix C presents a dendrogram that resulted from the cluster analysis, a summary of stations identified in each cluster, and a summary of chemical concentrations and toxicity responses. Results for the PCA include a screeplot of the variance explained by each principal component, loadings for the first five principal components, and scatter plots between principal components and biological endpoints (Appendix C).

4.3 SUMMARY

Overall conclusions from the exploratory analysis that helped inform the development of the predictive models included:

- The strong correlations indicate that there would be very little loss of information if sums of PAHs were used instead of individual analytes in the site-specific model development.
- Correlations between percent fines and other analyte concentrations in sediment were not particularly strong, with the exceptions of aluminum, selenium, ammonia, and TOC. However, for nearly all chemicals, there was an association between high chemical concentrations and high percent fines: high chemical concentrations occurred only in the sediments with the highest percent fines (but high percent fines did not always predict high chemical concentrations).
- Toxic responses are sometimes associated with high individual (or combined) chemical concentrations, but high chemical concentrations are not always associated with a toxic response.
- Diverse sources of chemicals in and the heterogeneity of the Study Area physical characteristics likely affected our ability to discern site-wide patterns.
- Some chemicals that have potential relationships with toxicity (e.g., TPH) have only a limited number of samples co-located with bioassays, which likely limited our ability to clearly assess this relationship.
- The range of responses among the various bioassays was limited and was skewed to low or no deleterious effects, which impaired identification of clear statistical relationships.

revealed substantial heterogeneity in the chemistry associated with the toxic responses.

Appendix C presents a dendrogram that resulted from the cluster analysis, a summary of stations identified in each cluster, and a summary of chemical concentrations and toxicity responses. Results for the PCA include a screeplot of the variance explained by each principal component, loadings for the first five principal components, and scatter plots between principal components and biological endpoints (Appendix C).

4.3 SUMMARY

Overall conclusions from the exploratory analysis that helped inform the development of the predictive models included:

- The strong correlations indicate that there would be very little loss of information if sums of PAHs were used instead of individual analytes in the site-specific model development.
- Correlations between percent fines and other analyte concentrations in sediment were not particularly strong, with the exceptions of aluminum, selenium, ammonia, and TOC. However, for nearly all chemicals, there was an association between high chemical concentrations and high percent fines: high chemical concentrations occurred only in the sediments with the highest percent fines (but high percent fines did not always predict high chemical concentrations).
- Toxic responses are sometimes associated with high individual (or combined) chemical concentrations, but high chemical concentrations are not always associated with a toxic response.
- Diverse sources of chemicals in and the heterogeneity of the Study Area physical characteristics likely affected our ability to discern site-wide patterns.
- Some chemicals that have potential relationships with toxicity (e.g., TPH) have only a limited number of samples co-located with bioassays, which likely limited our ability to clearly assess this relationship.
- The range of responses among the various bioassays was limited and was skewed to low or no deleterious effects, which impaired identification of clear statistical relationships.

5.0 DEVELOPMENT OF BENTHIC TOXICITY PREDICTION MODEL

Two principal models were chosen to determine if a predictive relationship between sediment chemistry and toxicity response could be developed for Portland Harbor. The two models, the FPM and LRM, were described in a previous technical memorandum (Windward 2005a). The models define the predictive relationship from different perspectives, although the goals of the models are similar: to develop a predictive relationship based on empirical data (i.e., sediment chemistry and toxicity test data) and to identify the principal chemical(s) that appear to define the relationship between sediment quality and toxicity.

The FPM focuses on identifying the chemicals that are apparently associated with observed toxicity and establishing SQVs for those chemicals based on minimizing errors (e.g., false positives and false negatives) and optimizing predictive reliability. The LRM focuses on developing mathematical models (using logistic regression) that describe the relationship between the probability of toxicity and chemical concentrations for each chemical. In addition to developing predictive mathematical models for individual chemicals, the LRM can also be used to combine multiple chemicals into a single logistic curve that provides a probability of toxicity for the chemical suite being considered.

In addition to the FPM and LRM, site-specific AETs were developed for Portland Harbor, and their reliability was evaluated. An AET is defined as the highest no-hit concentration of a given chemical. Above this threshold, all concentrations of that chemical are associated with a toxicity test endpoint response that is considered toxic (e.g., considered a hit based on the hit/no-hit definitions used for the data set).

This section presents a more complete explanation of each of these three methods, as well as an overview of how the models were applied to the Portland Harbor data set and the results of model development.

5.1 FLOATING PERCENTILE MODEL

The FPM was initially developed to improve the reliability of freshwater SQVs for Portland Harbor (ODEQ 1999) and Washington State (Ecology 2002, 2003). Unlike most other existing SQV sets, this model does not require the SQVs for all chemicals to be based on the same percentile of the hit or no-hit distribution. It is possible to minimize both false positive and false negative errors at the same time, as compared to other models, because the FPM is primarily eliminating prediction errors associated with the use of fixed percentiles to set SQVs for all chemicals. To date, FPM has been used in Washington State to develop SQVs for 11 metals, 16 individual PAHs, LPAHs, HPAHs, 4 phthalates, dibenzofuran, and total PCBs. These SQVs were derived using a large data set, primarily from western Washington and Oregon, including all of the Portland Harbor data that existed at that time (2001) and are currently applicable to freshwater sediments in Washington State (Ecology 2002).

The basic concept behind the FPM is for the user to select an optimal percentile of the data set that provides a low false negative rate and then adjust individual chemical concentrations upward until false positive rates are decreased to their lowest possible level while retaining the same low false negative rate. As shown in Figure 5-1, the y-axis is the percentile of each chemical's overall distribution and is not linearly related to toxicity. The green vertical line shows the concentration range within which toxicity does not occur, and the red vertical line shows the range within which toxicity occurs. These ranges may overlap due to site-specific or sample-specific variations in bioavailability or toxicity.

A constant percentile of the distribution that results in a low false negative rate is initially selected for all chemicals, represented by the blue dashed line. The difference between this constant percentile and the lower end of the toxicity range for each chemical is the area between the blue line and the red bar, and this is the source of most of the false positive errors.

The next step is to determine which chemicals are associated with false positive errors in the data set and adjust those concentrations upward until the lower ends of their toxicity ranges are reached (red bar). Above this point, false negatives will begin to increase. Above the red bar, both false negatives and false positives may occur, as is shown for Chemicals A, B, and C. This region is the range of concentrations over which sample-specific bioavailability plays an important role in toxicity, and therefore hit and no-hit samples are mixed together, causing both types of errors.

In Figure 5-1, Chemical B's concentration cannot be raised at all because it is already within its toxic concentration range. In any data set, a few chemicals will already be at a toxic level, giving rise to the low percentage of false negatives that the blue line represents. Some chemicals may show a sharper toxicity threshold (e.g., Chemical E). Others may not appear to be related to toxicity in the data set at all (e.g., Chemicals D and F). These chemical concentrations can be raised to their maximum percentile without any observed increase in toxicity. However, it may be safer in practice to raise them only to the point at which false positives no longer occur (represented by the green bar) or to similar thresholds such as AETs.

Once each chemical has been individually adjusted upward to the lower end of its toxicity range, the false positive rate will have been significantly reduced while the same low false negative rate is retained. Most chemicals should be at or near their actual toxicity range, rather than at a level arbitrarily assigned by a fixed percentile. In this manner, optimized site-specific SQVs can be developed for a number of different target false negative rates, allowing the trade-offs between false negatives and false positives to be evaluated and a final set of SQVs to be selected.

5.1.1 FPM Methodology

The modeling process for the FPM can be summarized in six steps as presented below. The first three steps, described in detail in Section 2.0, are identical to the data

organization steps used for the two other methods or models (AETs and LRM). Step 4 is also carried out for the AET model (see Section 5.2).

- **Step 1. Data Query** – The project database was queried to retrieve all of the chemistry and toxicity data for stations at which toxicity tests were conducted.
- **Step 2. Chemical Screening** – Analytes were screened out, as described in Section 2.2.2, based on the number of detected values, non-toxicity, and summation rules.
- **Step 3. Bioassay Statistical Analysis** – The toxicity results for each station were assigned a hit/no-hit status for each of the six endpoints (four individual and two pooled by species) and three effects levels.
- **Step 4. Creation of Hit and No-Hit Distributions** – The chemistry data for each analyte were then divided into hit and no-hit distributions and ranked in order of increasing concentration for each of the distributions.
- **Step 5. Development of Analyte Lists** – Analytes were evaluated using an analysis of variance (ANOVA) comparison of their hit and no-hit distributions to determine whether they were associated with toxicity. Analytes were retained for model development for each endpoint if they were associated with toxicity at two or three of the effects levels. Those chemicals for which the concentrations associated with bioassay hits versus no-hits could not be statistical distinguished were assigned values equivalent to AETs by the model.
- **Step 6. Selection of Optimal Chemical Concentrations** – Automated floating percentile macros and hand-optimization steps were used to identify chemical concentrations for each endpoint and effects level in order to minimize prediction errors.

As noted in Section 2.1.2, a minimum of 30 detected values was chosen as the lower limit for a chemical to be carried forward in the analyte list. Additional analytes, such as crustal elements, were also screened out prior to FPM model development (see Section 2.1.2). Chemicals retained for model development after the initial data organization and reduction are listed in Table 5-1.

Table 5-1. Analytes retained for FPM model development

Percent fines
4-Methylphenol
Aldrin
Ammonia
Antimony
Arsenic
Bis(2-ethylhexyl) phthalate
Butylbenzyl phthalate

Table 5-1. Analytes retained for FPM model development

Cadmium
Chromium
Copper
Dibutyltin
Dieldrin
Diesel-range hydrocarbons
Di-n-butyl phthalate
Hexachlorobenzene
alpha-Hexachlorocyclohexane
beta-Hexachlorocyclohexane
delta-Hexachlorocyclohexane
Lead
Mercury
Methoxychlor
Monobutyltin
Nickel
Pentachlorophenol
Phenol
Residual-range hydrocarbons
Silver
Sulfide
Tetrabutyltin
Total chlordane
Total DDTs
Total dioxins/furans
Total endosulfans
Total PAHs
Total PCBs
Tributyltin
Zinc

As part of Step 5, a second screening of the remaining data was conducted to remove chemicals that are not apparently associated with toxicity in this data set. This was accomplished by comparing the hit and no-hit distributions to determine if they were statistically different using an ANOVA comparison, $p \leq 0.05$. Experience with the application of the FPM has shown that chemicals with hit and no-hit distributions that are not statistically different do not affect the reliability of the SQVs developed using that data set. This was verified in some early runs on this project, as well as by recent projects conducted for the Washington State Department of Ecology (Ecology) (Avocet 2003), ODEQ (1999), San Francisco Bay (Germano & Associates 2004), and Los Angeles Harbor (unpublished).

Chemicals that are screened out at this stage would be assigned values equal to their AETs if they were retained in the model. Therefore, any analytes screened out by the

ANOVA test were assigned AET-equivalent values as part of the FPM SQV set. The development of site-specific AETs is further discussed in Section 5.2.

Table 5-2 presents the results of the ANOVA screening evaluation, which was initially conducted separately for each chemical, effects level, and endpoint combination. If a chemical showed a significant difference between the hit and no-hit distributions across two of the three effects levels, it was retained for that biological endpoint. In one of the exploratory runs for the FPM, chemicals were retained if there was a significant difference for any one of the effects levels, but the results indicated that these chemicals did not affect the reliability of the SQV set. As with the other chemicals that were screened out, the model assigned these chemicals their site-specific AETs as SQVs. Therefore, in the final run, only chemicals with significant differences for at least two of the effects levels were retained.

Certain chemicals had no significant differences for any of the hit/no-hit definitions or endpoints. These included: 4-methylphenol, aldrin, alpha- hexachlorocyclohexane, antimony, bis(2-ethylhexyl)phthalate, butylbenzyl phthalate, chromium, delta-hexachlorocyclohexane, dibutyltin, hexachlorobenzene, monobutyltin, pentachlorophenol, phenol, tetrabutyltin, total dioxins/furans, total endosulfans, and tributyltin. Additional chemicals that were not significant for each specific endpoint are shown in non-bolded text in Table 5-2.

Table 5-2. Chemical screening using ANOVA

Parameter	*Chironomus* Growth			*Chironomus* Mortality			*Chironomus* Pooled			*Hyalella* Growth			*Hyalella* Mortality			*Hyalella* Pooled		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
% Fines	1	1	0	1[a]	0	0	1[b]	1	0	1[b]	1[b]	1[b]	1[b]	0	0	1[b]	1[b]	1[b]
4-Methylphenol	1[a]	0	0	0	0	1	1[a]	0	0	0	0	0	0	0	0	0	0	0
Aldrin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
alpha-Hexachlorocyclohexane	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ammonia	1[b]	1[b]	1[a]	1[a]	1[a]	1	1[b]	1[b]	1[a]	1[b]	1[b]	0	0	1	1	1[b]	1[b]	1[a]
Antimony	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arsenic	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1
beta-Hexachlorocyclohexane	1[b]	1[b]	1[b]	1	1[b]	1[a]	1	1[a]	1[b]	0	0	0	1[b]	1[b]	1[b]	1	1[a]	1[a]
bis(2-Ethylhexyl) phthalate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Butylbenzyl phthalate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cadmium	1	1[b]	1[a]	1[a]	1	1	1[a]	1	1	1	0	0	0	1[a]	1[a]	1	1[a]	1
Chromium	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Copper	0	0	0	0	0	0	0	0	0	1[b]	1	0	0	0	0	1	1	1
delta-Hexachlorocyclohexane	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dibutyltin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dieldrin	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Diesel-range hydrocarbons	1[b]	1[b]	1[b]	1[b]	1[b]	1[b]	1[b]	1[b]	1[b]	0	0	1	1[b]	1[b]	1[b]	0	0	1[b]
Di-n-butyl phthalate	1[b]	1[b]	0	0	1	1	1[a]	1[b]	1	0	0	0	0	1	1	0	0	0
Hexachlorobenzene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lead	1	1[b]	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Mercury	1[b]	1[b]	1[b]	0	1	1	1[a]	1[b]	1[a]	1	0	0	1	1[a]	1[b]	1	1	1
Methoxychlor	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0
Monobutyltin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nickel	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1
Pentachlorophenol	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Phenol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Residual-range hydrocarbons	1[b]	1[b]	1[b]	1[a]	1[b]	1[b]	1[b]	1[b]	1[b]	0	0	0	1[b]	1[b]	1[b]	0	0	1[a]
Silver	1[a]	1[b]	1[b]	0	1	1[a]	1	1[a]	1[a]	0	0	0	1	1[b]	1[b]	1	1[a]	1[a]
Sulfide	1[b]	1[b]	1[b]	1[a]	1[b]	1	1[b]	1[b]	1[b]	0	0	0	1[b]	1[b]	1[a]	0	0	0

Table 5-2. Chemical screening using ANOVA

Parameter	*Chironomus* Growth			*Chironomus* Mortality			*Chironomus* Pooled			*Hyalella* Growth			*Hyalella* Mortality			*Hyalella* Pooled		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Tetrabutyltin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total chlordanes	**1**	**1**	**1**[a]	**0**	**1**	**1**	**0**	**1**	**1**	0	0	0	**1**	**1**	**1**[b]	0	0	0
Total DDTs	1	0	0	**1**	**1**[a]	**1**[b]	**0**	**1**[a]	**1**[a]	0	0	0	**1**[a]	**1**[b]	**1**[b]	0	0	0
Total dioxins/furans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total endosulfans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total PAHs	**1**[b]	**1**[b]	**1**[b]	**1**[a]	**1**[b]	**1**[b]	**1**	**1**[b]	**1**[b]	0	0	1	**1**[b]	**1**[b]	**1**[b]	0	0	1
Total PCBs	**1**	**1**[b]	**1**[b]	**0**	**1**	**1**[a]	**0**	**1**	**1**[a]	0	0	0	**1**[a]	**1**[b]	**1**[b]	0	0	0
Tributyltin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zinc	**1**	**1**[a]	**0**	0	0	0	1	0	0	**1**	**1**	**1**	0	0	0	**0**	**1**	**1**

0 – This chemical showed no apparent difference in its hit and no-hit distributions for this hit/no-hit definition
1 – This chemical showed significant differences in its hit and no-hit distributions for this hit/no-hit definition ($p < 0.05$)
Bold text and shading indicate that the chemical was retained for model development if statistical significance was observed at more than one effects level.
[a] $p < 0.005$
[b] $p < 0.0005$

The following chemicals had significant differences between their hit and no-hit distributions (in approximate order of greatest to least significance).

- ***Chironomus* Growth** – Diesel-range hydrocarbons, residual-range hydrocarbons, total PAHs, mercury, beta-hexachlorocyclohexane, sulfides, ammonia, silver, total PCBs, di-n-butyl phthalate, cadmium, lead, total chlordane, zinc, arsenic, percent fines
- ***Chironomus* Mortality** – Diesel-range hydrocarbons, residual-range hydrocarbons, total PAHs, beta- hexachlorocyclohexane, sulfides, total DDTs, ammonia, cadmium, silver, total PCBs, di-n-butyl phthalate, mercury, total chlordane
- ***Chironomus* Pooled** – Diesel-range hydrocarbons, residual-range hydrocarbons, sulfides, ammonia, total PAHs, mercury, beta-hexachlorocyclohexane, di-n-butyl phthalate, silver, total DDTs, percent fines, cadmium, total PCBs, total chlordane
- ***Hyalella* Growth** – Percent fines, ammonia, copper, arsenic, zinc, nickel, methoxychlor
- ***Hyalella* Mortality** – Diesel-range hydrocarbons, residual-range hydrocarbons, total PAHs, beta-hexachlorocyclohexane, sulfides, total PCBs, silver, mercury, total chlordane, cadmium, ammonia, di-n-butyl phthalate
- ***Hyalella* Pooled** – Percent fines, ammonia, beta-hexachloro-cyclohexane, silver, cadmium, arsenic, copper, mercury, nickel, zinc, methoxychlor

From the above lists and Table 5-2, it can be observed that petroleum-related analytes (PAHs, diesel-range hydrocarbons and residual-range hydrocarbons), as well as sulfides and ammonia, appear to be strongly associated with toxic responses for most endpoints and endpoint combinations. *Hyalella* growth is notably different from the other three individual endpoints, and the pooled *Hyalella* endpoint is strongly influenced by the *Hyalella* growth endpoint. *Hyalella* growth has its only strong correlations with percent fines and ammonia, has weaker correlations with various metals, and no correlation at all with petroleum analytes or most other organics. On the other hand, the two individual *Chironomus* endpoints respond very similarly to most chemicals and are also very similar to the *Chironomus* pooled endpoint.

It is also interesting to note that for most endpoints, bulk petroleum (diesel-range hydrocarbons and residual-range hydrocarbons) was somewhat more strongly correlated with toxicity than were total PAHs, in spite of the fact that PAHs were measured at all stations, and bulk petroleum was measured at only a subset of stations. This accords well with toxicological literature, which predicts that petroleum-based toxicity (narcosis) would be based on the total molecular concentrations of both aromatic and aliphatic constituents (Connell and Markwell 1992; Veith et al. 1983). Bulk petroleum

encompasses a greater percentage of the total constituents present than do total PAHs alone. However, both PAHs and bulk petroleum were retained for model development because bulk petroleum measurements are not available for many existing data sets or for all stations. This correlation indicates that in the future (i.e., during remedial design), it may be more appropriate to collect and use bulk petroleum data in comparison to SQVs than to use PAHs alone.

The last step in the modeling process, the selection of optimal chemical concentrations (Step 6), is particular to the FPM. The selection process occurs in two steps: an iterative automated step using an Excel® macro, and a hand-optimization step to address covariance and other issues that cannot be satisfactorily resolved by the macros alone. The Excel® macro uses the following approach to conduct the initial optimization:

- An appropriate incremental increase for testing is calculated for each analyte based on that analyte's complete concentration range (e.g., 1/10 of the difference between the highest and lowest concentration).
- The number of false positives contributed by each individual analyte is calculated, and the chemical contributing the most false positives is selected to begin the optimization procedure.
- The concentration for that analyte is increased by the chosen increment.
- After each incremental increase, false negative and false positive rates are recalculated for the entire SQV set.
- If the false negative rate increases, the chemical concentration is adjusted back down to its previous effects level, and that chemical is "locked in" at that level.
- If the false positive rate is reduced to zero, the chemical concentration is locked in at that effects level.
- If either of the above two conditions (i.e., Step 5 or 6) is met, the chemical is completed, and the macro moves on to the chemical with the next highest number of false positives. If neither condition is met, the macro raises the concentration by another increment and repeats Steps 4 to 7.
- Incremental increases and recalculations continue until every chemical has reached its toxicity threshold or a level at which it has no more false positives.

Through this process, it is possible to identify those analytes that have the greatest influence on toxicity in the data set (those with concentrations that cannot be increased without increasing false negatives) and those chemicals that have little or no influence on toxicity in the data set (those that can be increased to their highest concentrations with no effect on error rates).

An inspection of the results of the automated process, particularly when various starting percentiles are chosen, identifies analytes (often metals) with a high covariance in the

data set. It may also become apparent that other chemicals, such as PAHs, have relatively little effect individually but may act in an additive manner to cause toxicity. The automated process treats each chemical as an independent variable. If covariance or additive effects are pronounced in the data set, this can cause variation in the results depending on the starting values that are chosen. Covariance is present in most large data sets, including the Portland Harbor data set. This effect must be addressed through a final optimization step, requiring judgment on the part of the user to select the most appropriate values.

The spreadsheets used to develop the SQVs include a test macro and provide a place where candidate values can be entered and adjusted. The macro tests the results of each change with respect to all of the reliability parameters (this allows users to enter any SQVs of their choice and test their reliability against the regional data set). The following procedure is used for hand-optimization:

- To help minimize the effects of covariance, the values that result from the automated macro using various starting concentrations are scanned, and the lowest value for each chemical is selected as a starting point. These values are entered into the test area, and their reliability as a set is calculated.
- A false negative target of 5% is selected for the first optimization.
- The concentration of the chemical with the highest number of false positives is raised until either: 1) the false positives decrease to less than another chemical, or 2) the false negative target is reached. If the concentration of a chemical cannot be increased without exceeding the false negative target, it is locked in at that concentration.
- This process is repeated with each chemical in turn, always working with the chemical that has the most false positives remaining, until all chemicals are either locked in or have zero false positives remaining. This set of concentrations represents the recommended SQV set that corresponds to 5% false negatives.
- Next, the false negative target is raised in 5% increments to 10, 15, 20, and 25%, and the hand-optimization process is repeated for each false negative target, always building on the values already derived for the target below. This results in five sets of recommended SQVs for five target levels of false negatives.

Lastly, the SQVs are finalized by performing cross-checks. The following guidelines were followed in finalizing the SQVs:

- The resulting SQVs should be internally consistent within the same hit/no-hit definition. Specifically, chemical concentrations should increase or stay the same as the false negative rate increases and the false positive rate decreases. A range of 5 to 25% false negatives was used to

evaluate this guideline and to provide a range of options for selecting SQVs.

- The resulting SQVs should be consistent across different effects level definitions. Specifically, chemical concentrations should increase or stay the same as the adverse effects level increases. Effects Levels 1, 2, and 3 as previously defined were used in this process.
- The resulting SQVs should have equal or better reliability than those produced by the automated macros and all other available SQV sets.

Following each of these guidelines ensures that any anomalies produced by covariance or other interactions between chemicals in the data set are removed and addressed in a defensible manner.

5.1.2 Results of the FPM Runs

Tables 5-3, 5-4, and 5-5 present the results of the FPM for each of the 18 model runs and the selected analytes. Table 5-3 presents the comparative reliability for each of the 18 model runs against each of the seven reliability parameters selected for analysis (see Section 3.1.1 for definitions of the reliability parameters). Tables 5-4 and 5-5 present the proposed SQVs for conventionals/metals and organics, respectively, that resulted from these model runs. Backup spreadsheets that present the calculations in greater detail are included in Appendix D. As noted above, for each of the 18 model runs, 5 possible sets of SQVs were calculated based on a range of false negatives (i.e., 5, 10, 15, 20, and 25%) to provide an indication of trends in the modeling results and reliability parameters. By way of example, Tables 5-3, 5-4, and 5-5 present only one of these sets of results for each of the 18 model runs. In each case, the set of SQVs chosen was the one that had the most equal balance of false negatives and false positives, except that in no case was the level of false negatives allowed to increase above the 20% level. The following are notable from these results.

Reliability of individual and pooled endpoints

Ideally, both false negatives and false positives would be below 20%, and the overall reliability would be above 80%, the same goals used to select the Washington State freshwater standards using this model. In addition, predicted no-hit reliability would be above 90% in order to have greater confidence in defining a station as having no toxicity. In most cases, this was possible to achieve (exceptions are identified in Table 5-3).

Table 5-3. FPM reliability results

Endpoints by Biological Effects Level	Reliability Parameters						
	% False Negatives	% False Positives	% Sensitivity	% Efficiency	% Predicted Hit Reliability	% Predicted No-Hit Reliability	% Reliability
Level 1							
Chironomus growth	14	14	86	86	49	98	86
Chironomus mortality	19	**30**	81	70	42	93	**73**
Chironomus pooled	20	**35**	80	65	45	91	**69**
Hyalella growth	20	**53**	80	47	74	**55**	**69**
Hyalella mortality	20	23	80	78	35	96	78
Hyalella pooled	20	**34**	80	66	86	**56**	**76**
Level 2							
Chironomus growth	8	12	92	88	47	99	88
Chironomus mortality	21	21	79	79	39	96	79
Chironomus pooled	19	18	81	82	49	95	82
Hyalella growth	20	**51**	80	49	54	**76**	**62**
Hyalella mortality	10	8	90	92	53	99	92
Hyalella pooled	19	**45**	81	55	64	**74**	**68**
Level 3							
Chironomus growth	12	9	88	91	45	99	91
Chironomus mortality	20	18	80	82	35	97	82
Chironomus pooled	16	15	84	85	47	97	85
Hyalella growth	20	**54**	80	46	27	90	**53**
Hyalella mortality	11	7	89	93	53	99	93
Hyalella pooled	19	**44**	81	56	41	89	**63**

Bold text and shading identify exceptions.

Table 5-4. FPM SQVs – conventionals and metals

Endpoints by Biological Effects Level	Analytes										
	% Fines	Ammonia (mg/kg)	Sulfide (mg/kg)	Arsenic (mg/kg)	Cadmium (mg/kg)	Copper (mg/kg)	Lead (mg/kg)	Mercury (mg/kg)	Nickel (mg/kg)	Silver (mg/kg)	Zinc (mg/kg)
Level 1											
Chironomus growth	nc	180	32	24	3.6	562	nc	0.30	nc	1.8	nc
Chironomus mortality	nc	145	415	nc	1.5	nc	nc	0.73	nc	1.8	1,360
Chironomus pooled	88	165	115	22.9	1.5	562	178	0.63	nc	1.8	703
Hyalella growth	59	86	445	7.5	1.4	60	147	0.62	29	1.6	142
Hyalella mortality	--	335	60	nc	2.6	562	nc	0.73	nc	0.3	--
Hyalella pooled	57	103	29.1	7.5	1.5	350	147	0.14	65	1.8	740
Level 2											
Chironomus growth	nc	180	32	24	3.6	562	nc	0.63	nc	1.8	nc
Chironomus mortality	nc	170	415	nc	3.6	nc	nc	0.73	nc	1.8	nc
Chironomus pooled	nc	170	115	22.9	3.6	562	178	0.63	nc	1.8	nc
Hyalella growth	59	103	491	7.5	1.4	400	nc	nc	29	nc	142
Hyalella mortality	nc	335	415	nc	2.6	562	nc	0.73	nc	1.8	nc
Hyalella pooled	59	105	87.5	7.5	1.5	400	147	0.14	105	1.8	740
Level 3											
Chironomus growth	nc	280	415	34	3.6	562	nc	0.63	nc	1.8	nc
Chironomus mortality	nc	335	415	nc	3.6	nc	nc	0.73	nc	1.8	nc
Chironomus pooled	nc	280	415	nc	3.6	562	nc	0.63	nc	1.8	nc
Hyalella growth	62	103	nc	17.5	1.6	nc	nc	nc	29	nc	142
Hyalella mortality	nc	335	415	nc	2.6	562	nc	0.73	nc	1.8	nc
Hyalella pooled	62	105	nc	17.5	1.5	400	365	0.15	105	1.8	740

nc – FPM value could not be calculated because the chemical's toxicity threshold exceeds the maximum level found in the data set.

Table 5-5. FPM SQVs – organics and pesticides

Endpoints by Biological Effects Level	Analytes									
	beta-Hexachloro-cyclohexane (µg/kg)	Dieldrin (µg/kg)	Diesel-Range Hydrocarbons (mg/kg)	Di-n-butyl-phthalate (µg/kg)	Methoxyclor (µg/kg)	Residual Range Hydrocarbons (mg/kg)	Total Chlordane (calc'd) (µg/kg)	Total DDTs (calc'd) (µg/kg)	Total PAHs (calc'd) (µg/kg)	Total PCBs Aroclors (calc'd) (µg/kg)
Level 1										
Chironomus growth	9.6	37	340	420	nc	2,700	nc	11,500	1,270,000	3,500
Chironomus mortality	9.6	21.5	290	90	nc	2,700	nc	1,000	1,500,000	220
Chironomus pooled	8.9	9.28	290	65	nc	2,700	nc	220	22,000	300
Hyalella growth	20.3	0.907	14,000	1,000	6.8	17,000	32	nc	470,000	1,760
Hyalella mortality	9.6	21.5	300	90	nc	4,500	nc	12,900	1,500,000	4,400
Hyalella pooled	2.0	0.907	1,700	82	6.8	2,600	32	1,070	470,000	1,760
Level 2										
Chironomus growth	9.6	37	340	420	nc	4,500	nc	nc	1,270,000	3,500
Chironomus mortality	9.6	21.5	340	90	nc	2,700	nc	1,000	1,500,000	1,400
Chironomus pooled	9.6	21.5	340	90	nc	2,700	nc	1,000	1,500,000	1,400
Hyalella growth	20.3	9.28	14,000	1,000	10	--	nc	nc	2,110,000	2,310
Hyalella mortality	9.6	21.5	540	90	nc	10,000	nc	12,900	1,800,000	4,400
Hyalella pooled	2.5	1.45	4,700	450	10	10,000	32	1,250	1,710,000	2,310
Level 3										
Chironomus growth	9.6	37	340	nc	nc	4,500	nc	nc	1,270,000	3,500
Chironomus mortality	21	21.5	340	90	nc	4,500	nc	1,000	1,500,000	1,450
Chironomus pooled	9.6	21.5	340	90	nc	2,700	nc	1,000	1,500,000	1,400
Hyalella growth	20.3	21.5	14,000	nc	20	--	nc	nc	2,110,000	nc
Hyalella mortality	9.6	21.5	1,000	90	nc	10,000	nc	12,900	1,800,000	4,400
Hyalella pooled	2.5	21.5	4,700	450	20	10,000	67	11,500	1,710,000	3,370

nc – FPM value could not be calculated because the chemical's toxicity threshold exceeds the maximum level found in the data set.

As presented in Table 5-3, the most reliable endpoints at all effects levels were *Chironomus* growth and *Hyalella* mortality. The *Chironomus* mortality and *Chironomus* pooled endpoints also met the goals outlined above at Effects Levels 2 and 3. *Hyalella* growth had consistently poor reliability at all effects levels, and the *Hyalella* pooled endpoint was strongly affected by the *Hyalella* growth results and, thus, was also very unreliable. In the FPM model, pooling endpoints results in reliability values that tend toward the least reliable of the individual endpoints being pooled. Therefore, if one of the two endpoints being pooled is unreliable, the pooled endpoint generally is unreliable as well. This can also be seen in the *Chironomus* results, though the effect is less pronounced because both of the *Chironomus* endpoints have moderate to high reliability (see Table 5-3).

Comparison of FPM results to existing SQV sets and overall usability

A comparison of Table 5-3 to Tables 3-1 and 3-2 shows that the FPM results are a substantial improvement over any of the existing SQV sets, none of which are able to achieve the goals outlined above (see Section 3.2 for the results of the comparison of thee data set to existing SQVs). The quotient methods, though better than the existing SQV sets, were also unable to achieve this level of performance. At both Effects Level 2 and Level 3, the FPM results can be used to provide sets of SQVs with good overall reliability, low false negatives and false positives, and high predicted no-hit reliability. At Effects Level 1, the results are not ideal but still better than the existing alternatives evaluated in Section 3.0. Stations that exceed the proposed SQVs for Effects Levels 2 and 3 are presented in Figures 5-2 and 5-3, respectively. It should be noted that these figures show all surface sediment stations, both those with toxicity test results and those with chemistry alone.

Limited number of analytes associated with toxicity

As can be seen from Tables 5-4 and 5-5, there is a limited number of analytes for which FPM values can be calculated because the level at which these analytes reach their toxicity threshold is apparently above their concentration ranges in this data set. Alternatively, another chemical may covary with them and represent them in the SQV set. As will be seen in Section 5.2, this is also true when site-specific AETs are calculated. Although it is considered desirable to have as large an SQV set as possible, there is generally a limited number of chemicals associated with toxicity in Portland Harbor, and thus it is only possible to calculate the suite of SQVs presented in Tables 5-4 and 5-5.

Figures 5-4 and 5-5 show the locations of errors associated with the FPM at Effects Level 2 and Effects Level 3, respectively. At both of these levels, false negatives are rare and fairly randomly spread throughout the area. False positives are also scattered throughout the area, but there are a few clusters of false positives that are worth noting:

- Along the shore just southwest of RM 9
- Near Swan Island (around Portland Shipyard)

- South of RM 6 on the west bank of the river (offshore of Gasco)
- In the channel south of RM 5
- In the slip northeast of RM 4 (offshore of Schnitzer)
- On the east side of the river just south of RM 2

There are generally fewer false positives at Effects Level 3 than at Effects Level 2, as would be expected since Level 2 is more conservative. Areas with clusters of false positives may identify areas where chemical concentrations are high but toxicity does not occur because the chemicals are frequently in a form that is not bioavailable. In these areas, it may not be appropriate to rely on chemistry data alone and comparing them to SQVs. This disparity may be appropriate to address during design-level investigations for individual sediment management areas.

5.2 APPARENT EFFECTS THRESHOLDS

In addition to the FPM and LRM models, site-specific AETs were evaluated as a stand-alone SQV set, similar to the manner in which they have been used in marine areas of Washington State and in the Columbia River.

5.2.1 AET Methodology

The method used for the derivation of AETs is described in detail in Puget Sound Estuary Program (PSEP 1988), and the same general steps were followed for each of the six biological endpoints for each of the three effects levels as described below and illustrated in Figure 5-6. As noted earlier, the first three steps are identical to the data organization steps used for the FPM and LRM (see Section 2.0), and Step 4 is also performed for the FPM (see Section 5.1).

- **Step 1. Data Query** – The project database was queried to retrieve all of the chemistry and toxicity data for stations at which toxicity tests were conducted.
- **Step 2. Chemical Screening** – Analytes were screened out as described above, based on the number of detected values, non-toxicity, and summation rules.
- **Step 3. Bioassay Statistical Analysis** – The toxicity results for each station were assigned a hit/no-hit status for each of the six endpoints and three effects levels.
- **Step 4. Creation of Hit and No-Hit Distributions** – The chemistry data for each analyte were then divided into hit and no-hit distributions and ranked in order of increasing concentration for each of the distributions.
- **Step 5. Removal of Outliers** – The highest no-hit concentration was compared with the second highest no-hit concentration; and if the highest

was more than three times higher, it was designated as an outlier and removed from the no-hit distribution.

- **Step 6. Identification of AET** – The highest remaining no-hit concentration was designated as the AET. If the highest remaining no-hit concentration for an analyte was higher than the highest hit concentration, then a greater-than sign (>) was placed before the AET value to indicate that the actual AET may be higher than that value, or an AET may not exist for that chemical.

5.2.2 AET Results

AETs were calculated for all of the chemicals retained after the initial screening, as presented in Table 5-1. Tables 5-6, 5-7, and 5-8 show the AETs calculated for each endpoint and effects level. In addition, the lowest AETs (LAETs) and second-lowest AETs (2LAETs) for each effects level were also identified. The LAET and 2LAET have been used for regulatory purposes in Washington State to define the SQS and CSL under the cleanup program and to set dredging standards.

Chemicals or chemical summations are either reported at a fixed concentration or a concentration having a greater-than (>) sign preceding the concentration. Chemicals or chemical summations are considered to have usable AETs if they have at least one hit station with a concentration higher than the highest no-hit concentration (see Figure 5-6). Chemicals or chemical summations that have a greater-than symbol indicate that there was no hit station with that chemical at a higher concentration than the highest no-hit station. These values are not appropriate for use as SQVs.

Table 5-9 shows the reliability results for these AETs as a stand-alone SQV set. Results are similar to previous efforts conducted for Portland Harbor and the Columbia River. Although many of the reliability parameters show good performance, false negatives range from 60 to 90%, indicating that most of the stations exhibiting toxicity would not be identified by the AETs alone. The FPM was originally designed to correct this deficiency by reducing the false negative rates to below 20%.

Another notable feature of the results is that the *Hyalella* growth endpoint tends to not perform as well as other endpoints, having higher false negatives and/or lower reliability and predicted no-hit reliability. The LAET is also affected, inasmuch as many of its values are set by the *Hyalella* growth endpoint.

Table 5-6. Site-specific AETs – conventionals and metals

Endpoints by Biological Effects Levels	Analytes																
	% Fines	Ammonia (mg/kg)	Sulfide (mg/kg)	Antimony (mg/kg)	Arsenic (mg/kg)	Cadmium (mg/kg)	Chromium (mg/kg)	Copper (mg/kg)	Lead (mg/kg)	Mercury (mg/kg)	Nickel (mg/kg)	Silver (mg/kg)	Dibutyltin (µg/kg)	Mono-butyltin (µg/kg)	Tetra-butyltin (µg/kg)	Tributyltin (µg/kg)	Zinc (mg/kg)
Level 1																	
Chironomus growth	> 100	276	110	> 19.3	22.9	3.51	> 224	562	178	0.624	> 200	1.72	> 910	60	29	2,750	> 1,940
Chironomus mortality	> 100	276	110	> 19.3	> 34	1.42	> 224	> 1,080	> 1,290	0.722	> 200	1.72	840	110	> 97	> 2,750	1,360
Hyalella growth	89.46	242	445	1.36	16.9	1.42	> 224	348	147	0.624	53.2	1.63	120	100	2.8	430	703
Hyalella mortality	> 100	334	110	6.37	> 34	3.51	> 224	562	> 1,290	0.722	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
LAET	89.5	242	110	1.36	16.9	1.42	NA	348	147	0.624	53.2	1.63	120	60	2.8	430	703
2LAET	2A	276	110	6.37	22.9	1.42	2A	562	178	0.624	2A	1.72	840	100	29	2,750	1,360
Level 2																	
Chironomus growth	> 100	276	166	> 19.3	22.9	3.51	> 224	562	178	0.624	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
Chironomus mortality	> 100	276	110	> 19.3	> 34	3.51	> 224	> 1,080	> 1,290	0.722	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
Hyalella growth	98.4	> 352	491	6.37	16.9	1.42	> 224	400	> 1,290	> 2.01	102	> 4.44	320	100	9.3	460	703
Hyalella mortality	> 100	334	> 998	> 19.3	> 34	3.51	> 224	562	> 1,290	0.722	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
LAET	98.4	276	110	6.37	16.9	1.42	NA	400	178	0.624	102	1.72	320	100	9.3	460	703
2LAET	2A	276	166	2A	22.9	3.51	2A	562	2A	0.722	2A	1.72	2A	2A	2A	2A	2A
Level 3																	
Chironomus growth	> 100	276	166	> 19.3	> 34	3.51	> 224	562	> 1,290	0.624	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
Chironomus mortality	> 100	334	> 998	> 19.3	> 34	3.51	> 224	> 1,080	> 1,290	0.722	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
Hyalella growth	> 100	> 352	> 998	11.8	16.9	1.61	> 224	> 1,080	> 1,290	> 2.01	102	> 4.44	380	100	43	2,750	731
Hyalella mortality	> 100	334	> 998	> 19.3	> 34	3.51	> 224	> 562	> 1,290	0.722	> 200	1.72	> 910	> 110	> 97	> 2,750	> 1,940
LAET	NA	276	166	11.8	16.9	1.61	NA	562	NA	0.624	102	1.72	380	100	43	2,750	731
2LAET	2A	334	2A	2A	2A	3.51	2A	2A	2A	0.722	2A	1.72	2A	2A	2A	2A	2A

LAET – lowest apparent effects theshold
2LAET – second-lowest apparent effects threshold
NA – AETs could not be developed for any of the four endpoints.
2A – Fewer than two AETs could be developed among the four endpoints.
> Indicates that the true AET is unknown but greater than the value shown.

Table 5-7. Site-specific AETs – organics

Endpoints by Biological Effects Level	Analytes											
	Bis (2-ethylhexyl) phthalate (µg/kg)	Butylbenzyl phthalate (µg/kg)	Di-n-butyl phthalate (µg/kg)	Hexachloro-benzene (µg/kg)	4-Methyl-phenol (µg/kg)	Pentachloro-phenol (µg/kg)	Phenol (µg/kg)	Diesel-Range Hydrocarbons (mg/kg)	Residual-Range Hydrocarbons (mg/kg)	Total PAHs (calc'd) (µg/kg)	Total Dioxins/ Furans (calc'd) (pg/g)	Total PCBs Aroclors (calc'd) (µg/kg)
Level 1												
Chironomus growth	> 17,000	> 2,800	380	> 17.5	390	> 320	120	1,700	2,600	1,250,500	> 2,674.26	3,134
Chironomus mortality	9,800	1,200	170	> 17.5	> 510	> 320	120	1,700	2,600	1,250,500	> 2,674.26	3,365
Hyalella growth	3,000	240	1,000	16.8	> 510	19	22	14,000	17,000	470,060	2,399.087	1,760
Hyalella mortality	> 17,000	> 2,800	450	> 17.5	> 510	> 320	120	4,200	4,400	1,250,500	> 2,674.26	3,365
LAET	3,000	240	170	16.8	390	19	22	1,700	2,600	470,060	2,400	1,760
2LAET	9,800	1,200	380	2A	2A	2A	120	1,700	2,600	1,250,500	2A	3,134
Level 2												
Chironomus growth	> 17,000	> 2,800	1,000	> 17.5	> 510	> 320	120	4,200	4,400	1,250,500	> 2,674.26	3,134
Chironomus mortality	> 17,000	> 2,800	450	> 17.5	> 510	> 320	120	1,700	2,600	1,250,500	> 2,674.26	3,365
Hyalella growth	> 17,000	1,200	1,000	16.8	> 510	> 320	96	14,000	> 18,000	2,108,000	2,399.087	2,310
Hyalella mortality	> 17,000	> 2,800	450	> 17.5	> 510	> 320	120	4,700	10,000	1,708,600	> 2,674.26	3,365
LAET	NA	1,200	450	16.8	NA	NA	96	1,700	2,600	1,250,500	2,400	2,310
2LAET	2A	2A	450	2A	2A	2A	120	4,200	4,400	1,250,500	2A	3,134
Level 3												
Chironomus growth	> 17,000	> 2,800	> 1,800	> 17.5	> 510	> 320	120	4,200	4,400	1,250,500	> 2,674.26	3,365
Chironomus mortality	> 17,000	> 2,800	450	> 17.5	> 510	> 320	120	1,700	3,600	1,250,500	> 2,674.26	3,365
Hyalella growth	> 17,000	1,200	> 1,800	> 17.5	> 510	> 320	96	14,000	> 18,000	2,108,000	> 2,674.26	> 3,365
Hyalella mortality	> 17,000	> 2,800	450	> 17.5	> 510	> 320	120	4,700	10,000	1,708,600	> 2,674.26	3,365
LAET	NA	1,200	450	NA	NA	NA	96	1,700	3,600	1,250,500	NA	3,365
2LAET	2A	2A	450	2A	2A	2A	120	4,200	4,400	1,250,500	2A	3,365

LAET – lowest apparent effects threshold
2LAET – second-lowest apparent effects threshold
NA – AETs could not be developed for any of the four endpoints.
2A – Fewer than two AETs could be developed among the four endpoints.
> Indicates that the true AET is unknown but greater than the value shown.

Table 5-8. Site-specific AETs – pesticides

Endpoints by Biological Effects Level	Analytes								
	Aldrin (µg/kg)	Dieldrin (µg/kg)	alpha-Hexachloro-cyclohexane (µg/kg)	beta-Hexachloro-cyclohexane (µg/kg)	delta-Hexachloro-cyclohexane (µg/kg)	Methoxychlor (µg/kg)	Total Chlordane (calc'd) (µg/kg)	Total DDTs (calc'd) (µg/kg)	Total Endosulfan (calc'd) (µg/kg)
Level 1									
Chironomus growth	25.9	9.28	2.98	8.5	> 1.26	> 19.8	55.46	11,480	3.41
Chironomus mortality	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	11,480	0.943
Hyalella growth	10.6	0.907	0.812	20.3	0.965	6.18	32.2	> 16,170.5	3.14
Hyalella mortality	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	11,480	21.1
LAET	10.6	0.907	0.812	8.5	0.965	6.18	32.2	11,480	0.943
2LAET	25.9	9.28	2.98	9.56	2A	2A	55.46	11,480	3.14
Level 2									
Chironomus growth	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	> 16,170.5	13.5
Chironomus mortality	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	11,480	3.41
Hyalella growth	10.6	9.28	2.89	20.3	> 1.26	9.98	> 668.8	> 16,170.5	13.6
Hyalella mortality	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	11,480	21.1
LAET	10.6	9.28	2.89	9.56	NA	9.98	67.42	11,480	3.41
2LAET	30	21.5	2.98	9.56	2A	2A	67.42	11,480	13.5
Level 3									
Chironomus growth	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	> 16,170.5	13.5
Chironomus mortality	30	21.5	2.98	20.3	> 1.26	> 19.8	67.42	11,480	21.1
Hyalella growth	> 30	> 21.5	2.89	20.3	> 1.26	> 19.8	> 668.8	> 16,170.5	21.1
Hyalella mortality	30	21.5	2.98	9.56	> 1.26	> 19.8	67.42	11,480	21.1
LAET	30	21.5	2.89	9.56	NA	NA	67.42	11,480	13.5
2LAET	30	21.5	2.98	9.56	2A	2A	67.42	11,480	21.1

LAET – lowest apparent effects threshold
2LAET – second-lowest apparent effects threshold
NA – AETs could not be developed for any of the four endpoints.
2A – Fewer than two AETs could be developed among the four endpoints.
> Indicates that the true AET is unknown but greater than the value shown.

Table 5-9. Reliability of site-specific AETs

Endpoints by Biological Effects Level	% False Negatives	% False Positives	% Sensitivity	% Efficiency	% Predicted Hit Reliability	% Predicted No-Hit Reliability	% Reliability
Level 1							
Chironomus growth	66	1	34	99	91	91	91
Chironomus mortality	74	1	26	99	86	83	83
Hyalella growth	67	4	33	96	94	43	55
Hyalella mortality	73	2	27	98	67	90	89
LAET	65	4	35	96	97	30	48
2LAET	73	3	27	97	74	81	80
Level 2							
Chironomus growth	71	0	29	100	88	92	92
Chironomus mortality	68	1	32	99	92	90	90
Hyalella growth	85	2	15	98	88	61	63
Hyalella mortality	60	0	40	100	89	95	94
LAET	79	1	21	99	96	51	56
2LAET	68	1	32	99	92	90	90
Level 3							
Chironomus growth	65	0	35	100	86	95	95
Chironomus mortality	64	0	36	100	90	93	93
Hyalella growth	89	1	11	99	83	82	82
Hyalella mortality	61	0	39	100	88	95	95
LAET	77	1	23	99	95	72	74
2LAET	64	0	36	100	89	94	94

LAET – lowest apparent effects threshold
2LAET – second-lowest apparent effects threshold

5.3 LOGISTIC REGRESSION ANALYSIS

The LRM approach was first proposed in 1999 as an alternative to threshold methods used for developing SQVs (Field et al. 1999; Field et al. 2002). A large national data set consisting of over 3,000 marine/estuarine sediment samples with matched chemistry and toxicity test results (two species of marine/estuarine amphipods) was assembled. On a study-by-study basis, the data were screened into three categories for each selected analyte: 1) non-toxic samples, 2) toxic samples with a chemical concentration greater than the mean concentration in the non-toxic samples, and 3) toxic samples with a chemical concentration lower than the mean concentration in the non-toxic samples. The designation as toxic was based on a statistically significant difference from the negative control and survival less than 90% (i.e., the minimum acceptable control survival). In this application of the LRM approach, the designation as toxic was based on 90% difference from control (Effects Level 1) plus two additional effects levels (Effects Levels 2 and 3), described earlier.

5.3.1 LRM Methodology

Following the general approach presented by EPA (EPA 2005b), LRMs were developed for the Portland Harbor data set. The steps of the modeling process are briefly described below. The first three general steps are the same as those used for both the FPM and for deriving site-specific AETs.

- **Step 1. Data Query** – The project database was queried to retrieve all of the chemistry and toxicity data for stations at which toxicity tests were conducted.
- **Step 2. Chemical Screening** – Analytes were screened out as described below, based on the number of detected values and summation rules.
- **Step 3. Bioassay Statistical Analysis** – The toxicity results for each station were assigned a hit/no-hit status for each of the three endpoints and three effects level definitions (see below).
- **Step 4. Chemistry and Toxicity Data** – Toxic stations that had concentrations less than the mean concentration for the non-toxic stations were identified. The set of data excluding these low concentration toxic stations constituted the "screened data set" upon which the logistic regression model for this chemical is based.
- **Step 5. Logistic Regression Model** – A logistic regression model using the screened data set relating toxicity to $\log_{10}$ concentration was applied. This resulted in a model of the following form for each analyte:

$$p = \frac{\exp(B_0 + B_1(x))}{1 + \exp(B_0 + B_1(x))} \qquad \text{Equation 1}$$

where:

p = probability of observing a toxic effect based on a single chemical (x)
x = $\log_{10}$ chemical concentration
B_0 = intercept parameter
B_1 = slope parameter

- **Step 6. Model Information** – Goodness of fit and other information useful in assessing the model were compiled (i.e., total samples and number of toxic samples retained in the screened data set; Chi-square statistics, likelihood ratio R^2 or R^2_L (Menard 2000), and concentration interval plots showing the data with the best-fit model).
- **Step 7. Repeat Model Process** – For a given biological endpoint, Steps 4 through 6 were repeated for every individual chemical analyte.
- **Step 8. Model Assessment** – Models with poor fit or insufficient data were excluded from further consideration. These were models that had Chi-square p-values greater than 0.01 or had zero or one hit retained in the screened data set. Models that had low R^2_L (< 0.20, an arbitrary cutoff) or had fewer than five hits retained in the screened data set were flagged as being unreliable but were retained in the multi-chemical modeling process (Step 9).
- **Step 9. Multi-Chemical Model Construction** – A multi-chemical model was constructed to predict the probability of a toxic effect from the mixture of contaminants observed in a sample. Each sample had a set of concentrations for the full suite of chemical results reported for that sample, and each of these concentrations had an associated probability of toxicity (p) predicted from the individual chemical models constructed in Step 5. The maximum value among these individual predictions of toxicity was used as the single best prediction of a toxic effect for each sample. The multi-chemical model related this maximum probability of a toxic effect (max_p) to the observed toxicity for the full set of site data (i.e., samples that had been screened out in Step 4 are included here). This was essentially a calibration step to accommodate the screened-out data and to produce a relative probability of toxicity that was as accurate as possible for the full set of data. Just as Equation 1 predicted the probability of toxicity as a linear function of chemical concentration on the logistic scale, the multi-chemical model (Equation 2) predicted the overall probability of toxicity as a linear function of max_p on the logistic scale:

$$PrMax = \frac{\exp(b_0 + b_1(max_p))}{1 + \exp(b_0 + b_1(max_p))} \qquad \text{Equation 2}$$

where:

PrMax = overall predicted probability of a toxic effect for a sample, based on all chemicals present in that sample

max_p = maximum predicted probability of toxicity across all analytes (maximum p for all individual chemical models constructed in Step 5)

b_0 = intercept parameter

b_1 = slope parameter

The accuracy of the PrMax predictions of a toxic effect for each biological endpoint are discussed in the results section (Section 5.3.2).

The chemical screening in Step 2 used a minimum of 30 detected values as the lower limit for inclusion on the analyte list (see Section 2.1.2), similar to that used for the FPM and AETs. Many analytes were not detected in Portland Harbor or were detected in very few locations. Many of these chemicals are represented in the final model outcome as part of a sum. The LRM approach is not adversely affected by multi-collinearity (i.e., correlation among chemical endpoints). Consequently, LRMs were built for some individual analytes that comprise sums (e.g., individual PAHs) in addition to the sums to which they contribute. This approach was taken to provide site-specific predictions of toxicity for as many target analytes as possible. In addition, percent fines, bulk sediment ammonia and sulfides were also retained in the analysis because of their apparently strong correlations with toxicity in some biological endpoints. Chemicals used in the LRM development are listed in Table 5-10.

Table 5-10. Analytes included in the set of initial individual LRMs

Conventionals
Ammonia
Percent fines
Sulfide
Dioxins/Furans
1,2,3,7,8-Pentachlorodibenzofuran
Pentachlorodibenzo-p-dioxin homologs
TEQ mammal (0.5 detection limit)
Total dioxins/furans
Metals
Aluminum
Antimony
Arsenic
Cadmium
Chromium
Copper
Lead

Table 5-10. Analytes included in the set of initial individual LRMs

Mercury
Nickel
Selenium
Silver
Zinc
Organotins
Butyltin
Dibutyltin
Tributyltin
Pesticides and PCBs
Aldrin
alpha-Hexachlorocyclohexane
beta-Hexachlorocyclohexane
delta-Hexachlorocyclohexane
Carbazole
Methoxychlor
cis-Nonachlor
trans-Nonachlor
Total chlordane
Total DDD
Total DDE
Total DDT
Total endosulfan
Total PCBs
PAHs
2-Methylnaphthalene
Acenaphthene
Acenaphthylene
Anthracene
Benzo(a)anthracene
Benzo(a)pyrene
Benzo(b)fluoranthene
Benzo(ghi)perylene
Benzo(k)fluoranthene
Chrysene
Dibenzanthracene
Dibenzofuran
Fluoranthene
Fluorene
Indeno(c,d)pyrene
Naphthalene
Phenanthrene
Pyrene
Total LPAH
Total HPAH
Total PAH

Table 5-10. Analytes included in the set of initial individual LRMs

Phenols and Phthalates
4-Methylphenol
Bis(2-ethylhexyl)phthalate
Butylbenzylphthalate
Dibutylphthalate
Pentachlorophenol
Phenol
Other Organics
Diesel-range hydrocarbons
Hexachlorobenzene
Residual-range hydrocarbons

Details regarding individual analyte selection by chemical group are provided below.

- **Dioxins/Furans** – Correlations were high among individual dioxin/furan isomers, homologs, and totals, with a few exceptions. Several individual furans and a dioxin homolog had substantial variation in the correlation with total dioxins/furans. Correlations were high among these individual furans, and among the homologs, so only one endpoint from each was retained in the LRM process. Total dioxins/furans, plus 1,2,3,7,8-pentachloro-dibenzofuran and pentachlorodibenzo-p-dioxin homologs; and one TEQ (total dioxin/furan TEQ for mammals with non-detects at 0.5 detection limit) were retained.
- **DDTs** – Correlations between total DDTs and the individual isomers were good, though better for 44-DDD than the others. From a toxicological standpoint, it may be worthwhile to have separate SQVs for the intermediate sums. Total DDD, total DDE, total DDT, and the sum total DDTs were retained.
- **Organotins (as ions)** – Correlations were high between tetra- and tributyltin; also between mono- and dibutyltin. Tetrabutyltin has fewer detected values but correlates quite well with tributyltin. Monobutyltin, dibutyltin, and tributyltin were retained.
- **Pesticides** – Linear correlations among total chlordane and the chlordane and nonachlor endpoints were good, with the exception of nonachlor (cis- and trans-). Total endosulfan; hexachlorocyclohexane (alpha-, beta-, and delta-); nonachlor (cis- and trans); and total chlordane were retained.
- **PAHs** – Individual PAHs were highly correlated with their respective sums (total PAHs and HPAHs). Due to the particular interest in the PAHs, individual PAHs, plus total LPAHs, total HPAHs, total PAHs, diesel-range hydrocarbons, and residual-range hydrocarbons were retained.
- **Metals and Crustal Elements** – All individual analytes, including selenium and aluminum, were retained.

- **Conventionals** – Because of observed correlations with some biological endpoints, bulk sediment sulfides, ammonia, and percent fines were retained.

For the statistical analysis of the toxicity data (Step 3) there were 21 possible biological endpoints (four individual endpoints plus three pooled endpoints, each at three different effects levels). Discussion with the EPA and its partners indicated that they were primarily interested in either a pooled species endpoint or in mortality but not in the growth endpoint alone (EPA 2005a). Consequently, the LRM approach was run on the *Chironomus* pooled endpoint (growth and mortality combined) and the *Hyalella* pooled endpoint (growth and mortality combined). Concern regarding the *Hyalella* growth endpoint (see Section 6.1) resulted in a third set of runs for the *Hyalella* mortality endpoint alone. Each of these three endpoints was run for each of the three effects levels, resulting in nine different biological endpoints. These are summarized in Table 5-11. A pooled species endpoint is a hit when either the growth or mortality endpoint was a hit.

Table 5-11. Hits for biological endpoints used in the LRM

Effects Level	Number of Biological Hits (percent)[a]		
	***Chironomus* pooled**	***Hyalella* pooled**	***Hyalella* mortality**
Level 1	56 (26%) [16]	158 (73%) [16]	30 (13%) [3]
Level 2	42 (18%) [0]	116 (50%) [0]	20 (9%) [0]
Level 3	32 (14%) [0]	64 (27%) [0]	18 (8%) [0]

a The denominator used to determine the percentage of hits excludes the number of statistically indeterminate samples shown in brackets.

5.3.2 Results of the LRM Runs

LRMs were developed for each of the chemical analytes identified in Table 5-10 and the biological endpoints identified in Table 5-11. The results for the individual chemical models are presented in Appendix E and by way of example in Figure 5-7. The nine models constructed for each chemical analyte are shown on a single page. For each plot within a page, the $\log_{10}$ chemical concentration is shown on the x-axis and the proportion of samples toxic within a concentration interval are shown on the y-axis. The symbol plotted at each (x,y) value is the number of samples within that concentration interval. All biological endpoints for an effects level are shown on a single row of plots, and all endpoints for a species are shown in a single column of plots. The title of each plot indicates the biological endpoint (e.g., hym.80 is Effects Level 2 [80% difference] for *Hyalella* mortality). Several items should be noted when interpreting the results presented in Figure 5-7 and Appendix E table and figures:

- Some of the models maintained a very low probability of toxicity throughout the range of concentrations observed, indicated by curves that ended with y-values less than 0.5 (e.g., hym.70 in Figure 5-7). These are chemicals that do not consistently result in high probabilities of toxicity within this data set. Other models reached a level of 100% toxicity within the observed range (e.g., hypool.90 in Figure 5-7), indicating a strong correlation between chemistry and toxicity within the screened data set.
- Samples with high concentrations and no toxicity can be observed as points falling well below the LRM line (e.g., Appendix E, Figure E-5 for antimony, for all endpoints except hypool.90).
- In general, endpoints with high base toxicity rates (e.g., the Effects Level 1 *Hyalella* pooled endpoint) tend to suggest a better relationship between chemistry and probability of toxicity because of the larger number of toxic stations available to define the curve.
- Chemicals that have very few toxic stations retained in the screened data set are ones in which the concentrations for toxic and non-toxic stations are not very different (see Appendix E, Table E-1, for individual LRMs).

The PrMax predictions for each sample are compared to the actual observed toxicity for the entire data set (i.e., it includes predictions for the toxic samples that were excluded from the screened data set in Step 4, as described in Section 5.2.1). These were done both as graphic and tabular comparisons.

Graphical comparisons

The PrMax predictions were plotted against the observed probability of toxicity using the observed toxic/non-toxic samples grouped by PrMax values in intervals of 0.05 (e.g., 0 to 0.05, 0.05 to 0.10). The plots show the median PrMax value among the grouped data (this may not be midpoint of the interval bounds but is usually close) vs. the ratio of toxic samples among the binned data. At each point, the number of samples in the bin is shown. Some PrMax intervals may be empty. The 1:1 line is shown on the graph for reference. Accurate predictions by the PrMax model will place the data points close to this line throughout the range. Figure 5-8 presents the data for each of the three biological endpoints at three effects levels vs. their PrMax values.

Tabular comparisons

Table 5-12 shows the predicted and observed levels of toxicity in five PrMax categories: < 20%, 20 to 40%, 40 to 60%, 60 to 80% and > 80%. Each sample had a PrMax value calculated from the chemical concentrations and an observed toxicity status. The PrMax value determined which column of the table the sample fell into, and its toxicity status determined in which row of that column the sample was placed. Once all samples had been placed in one column and one row, the percent toxic for each column was computed and compared to the average of the predicted toxicity (mean PrMax values) for all samples in that column. These tables identified the number and

type of errors (e.g., false positives are the non-toxic samples with high PrMax values, and false negatives are the toxic samples with low PrMax values), how the samples were distributed with respect to both observed toxicity status and chemistry, and the relationship between the observed toxicity and the relative predicted toxicity value (PrMax) derived from the chemical concentrations.

Table 5-12. Observed vs. predicted probabilities of toxicity

		Probability of Toxicity					
		< 20%	**20 – 40%**	**40 – 60%**	**60 – 80%**	**> 80%**	**Total**
PrMax for *Chironomus* Pooled Level 1 (90%)[b]							
Predicted	mean	11%	28%	51%	62%	NA	
Observed	non-toxic[a]	92	49	17	3	0	161
	toxic[a]	14	12	14	16	0	56
	total[a]	106	61	31	19	0	217
	% toxic	13%	20%	45%	84%	NA	
Difference (predicted mean vs. observed percent toxic)		-2%	8%	6%	-22%		
PrMax for *Chironomus* Pooled Level 2 (80%)							
Predicted	mean	7%	30%	51%	66%	NA	
Observed	non-toxic[a]	154	23	10	4	0	191
	toxic[a]	14	5	7	16	0	42
	totals[a]	168	28	17	20	0	233
	% toxic	8%	18%	41%	80%	NA	
Difference (predicted mean vs. observed percent toxic)		-1%	12%	10%	-14%		
PrMax for *Chironomus* Pooled Level 3 (70%)							
Predicted	mean:	6%	28%	53%	64%	NA	
Observed	non-toxic[a]	171	23	2	5	0	201
	toxic[a]	12	4	5	11	0	32
	totals[a]	183	27	7	16	0	233
	% toxic	7%	15%	71%	69%	NA	
Difference (predicted mean vs. observed percent toxic)		-1%	13%	-18%	-5%		
PrMax for *Hyalella* Pooled Level 1 (90%)[b]							
Predicted	mean	14%	29%	51%	76%	81%	
Observed	non-toxic[a]	2	6	10	31	10	59
	toxic[a]	0	5	6	80	67	158
	totals[a]	2	11	16	111	77	217
	% toxic	0%	45%	38%	72%	87%	
Difference (predicted mean vs. observed percent toxic)		14%	-16%	13%	4%	-6%	

Table 5-12. Observed vs. predicted probabilities of toxicity

		Probability of Toxicity					
		< 20%	20 – 40%	40 – 60%	60 – 80%	> 80%	Total
PrMax for *Hyalella* Pooled Level 2 (80%)							
Predicted	mean	16%	29%	52%	66%	NA	
Observed	non-toxic[a]	19	26	44	28	0	117
	toxic[a]	3	15	37	61	0	116
	totals[a]	22	41	81	89	0	233
	% toxic	14%	37%	46%	69%	NA	
Difference (predicted mean vs. observed percent toxic)		2%	-8%	6%	-3%		
PrMax for *Hyalella* Pooled Level 3 (70%)							
Predicted	Mean	13%	30%	50%	63%	NA	
Observed	non-toxic[a]	86	60	22	1	0	169
	toxic[a]	12	25	19	8	0	64
	totals[a]	98	85	41	9	0	233
	% toxic	12%	29%	46%	89%	NA	
Difference (predicted mean vs. observed percent toxic)		1%	1%	4%	-26%		
PrMax for *Hyalella* Mortality Level 1 (90%)[c]							
Predicted	mean	8%	28%	52%	63%	NA	
Observed	non-toxic[a]	179	14	6	1	0	200
	toxic[a]	16	1	6	7	0	30
	totals[a]	195	15	12	8	0	230
	% toxic	8%	7%	50%	88%	NA	
Difference (predicted mean vs. observed percent toxic)		0%	21%	2%	-25%		
PrMax for *Hyalella* Mortality Level 2 (80%)							
Predicted	Mean	3%	22%	47%	71%	NA	
Observed	non-toxic[a]	199	6	5	3	0	213
	toxic[a]	7	1	4	8	0	20
	totals[a]	206	7	9	11	0	233
	% toxic	3%	14%	44%	73%	NA	
Difference (predicted mean vs. observed percent toxic)		0%	8%	3%	-2%		
PrMax for *Hyalella* Mortality Level 3 (70%)							
Predicted	mean	3%	34%	49%	70%	NA	
Observed	non-toxic[a]	204	5	4	2	0	215
	toxic[a]	7	2	1	8	0	18
	totals[a]	211	7	5	10	0	233
	% toxic	3%	29%	20%	80%	NA	
Difference (predicted mean vs. observed percent toxic)		0%	5%	29%	-10%		

[a] Number of samples.
[b] Sixteen indeterminate samples were excluded from analysis.
[c] Three indeterminate samples were excluded from analysis.

The following observations can be made from the figures and tables.

- ***Chironomus* pooled endpoints** – For Level 1, the PrMax tends to overestimate toxicity for values between 0.25 and 0.55, as indicated by the curve of the data below the 1:1 line (Figure 5-8). It also underestimates toxicity at values greater than 0.6 (the data extend above the 1:1 line in Figure 5-8, and the difference between predicted and observed probabilities of toxicity in Table 5-12 are negative). At the higher effects levels, similar patterns are observed. PrMax predictions at Level 2 provide a fairly good fit to the data, with the exception of some overestimations for PrMax values less than 0.5 and underestimations for PrMax values exceeding 0.65.
- ***Hyalella* pooled endpoints** – For Level 1, most of the samples have PrMax values greater than 0.7. The PrMax predictions are fairly accurate in this region, with differences between observed and predicted toxicities less than 10%. There are a few false negatives (observations far above the 1:1 line for lower PrMax values, Figure 5-8). Under this effects level and biological endpoint, 73% of the samples are toxic, and most of them are predicted to have high probabilities of toxicity by their PrMax values (147 of the 158 toxic samples have PrMax values > 0.6, Table 5-12). Observed toxicity for the higher effects levels (Levels 2 and 3) match their PrMax predictions fairly well.
- ***Hyalella* mortality endpoints** – Very few samples are considered toxic for these endpoints, with a base toxicity rate ranging from 13% (Level 1) to 8% (Level 3). As a result of the fact that there were very few toxic samples in the data set, the predictions of toxicity from the PrMax values tend to be lower, which coincides with the lower observed toxicity. This results in pretty good non-toxic reliability, but the few toxic samples are poorly predicted with low PrMax values.

If the data distributions for the toxic and non-toxic samples overlap substantially, then the samples that were screened out during the initial individual chemical model fitting (Step 4) will reduce the accuracy at the low end of the predicted probability scale in this assessment phase: they will be toxic stations with low max_p values (i.e., false negatives). As a result, the PrMax value will be scaled down to accommodate these screened-out stations. This phenomenon was observed for nearly all endpoints, as indicated by the presence of toxic samples in all regions of the PrMax range (Table 5-12).

Identification of the optimal toxicity threshold

The PrMax threshold that is used to predict toxic stations can be set at any point within the PrMax range of zero to one. The seven reliability parameters (Section 3.1.1 and Figure 3-1) were computed for PrMax thresholds between zero and one, at intervals of 0.01 (i.e., 0.01, 0.02, 0.03, ... 0.98, 0.99). At each threshold, a station with a PrMax

value greater than the threshold is a predicted hit, and a station with a PrMax value at or below the threshold is a predicted no-hit. All seven parameters are shown on a single graph for increasing PrMax thresholds. A graph is displayed for each of the three biological endpoints (*Hyalella* mortality, *Hyalella* pooled, and *Chironomus* pooled) at the three effects levels (Figure 5-9).

Selection of a threshold can be based on an assessment of the error rates and the overall reliability. A lower threshold will identify more stations as potentially toxic, resulting in higher sensitivity but at the expense of lower efficiency and higher false positives. Using the same targets outlined for the FPM, a threshold that provided both false negative and false positive error rates below 20% and an overall reliability above 80% was sought. Unfortunately, this was unattainable with these data.

If there is substantial overlap in concentrations for the toxic and non-toxic distributions, then false positives and false negatives are closely tied, and false positives cannot be reduced without increasing false negatives, or vice versa. The optimal threshold is identified as the point where false positives and false negatives are jointly optimized (i.e., where the two lines cross in the graphs, see Figure 5-9). If one of the error rates could be improved at very little loss to the other, then the threshold could be adjusted to maximize overall reliability. Alternatively, an *a priori* threshold of PrMax > 0.6 could be selected.

Reliability Results

Reliability parameters are presented in Table 5-13 for a PrMax threshold of 0.6; the error-optimized threshold, as described above, is indicated by the shaded rows. The threshold of 0.6 has good accuracy for predicting toxicity (false positives are low, and efficiency is high) for all endpoints, except for the *Hyalella* pooled endpoint (Levels 1 and 2). Reliability results for the error optimization threshold show that both errors cannot be simultaneously maintained below a reasonable level (approximately 20%), except for *Hyalella* mortality at Levels 2 and 3.

Table 5-13. Reliability parameters for optional toxicity thresholds for all endpoints

Endpoint[a]	PrMax Threshold	% False Negatives	% False Positives	% Sensitivity	% Efficiency	% Predicted Hit Reliability	% Predicted No-Hit Reliability	% Reliability
Level 1								
Chironomus pooled	**0.23**	**32%**	**34%**	**68%**	**66%**	**41%**	**85%**	**66%**
	0.60	71%	2%	29%	98%	84%	80%	80%
Hyalella mortality	**0.12**	**37%**	**21%**	**63%**	**79%**	**31%**	**93%**	**77%**
	0.60	77%	1%	23%	100%	88%	90%	90%
Hyalella pooled	0.60	7%	69%	93%	31%	78%	62%	76%
	0.78	**29%**	**39%**	**71%**	**61%**	**83%**	**44%**	**68%**
Level 2								
Chironomus pooled	**0.15**	**26%**	**26%**	**74%**	**74%**	**38%**	**93%**	**74%**
	0.60	62%	2%	38%	98%	80%	88%	87%
Hyalella mortality	**0.08**	**15%**	**15%**	**85%**	**85%**	**35%**	**98%**	**85%**
	0.60	60%	1%	40%	99%	73%	95%	94%

Table 5-13. Reliability parameters for optional toxicity thresholds for all endpoints

Endpoint[a]	PrMax Threshold	% False Negatives	% False Positives	% Sensitivity	% Efficiency	% Predicted Hit Reliability	% Predicted No-Hit Reliability	% Reliability
Hyalella pooled	**0.55**	**34%**	**36%**	**66%**	**64%**	**64%**	**65%**	**65%**
	0.60	47%	24%	53%	76%	69%	62%	64%
Level 3								
Chironomus pooled	**0.10**	**25%**	**26%**	**75%**	**74%**	**32%**	**95%**	**74%**
	0.60	66%	2%	34%	98%	69%	90%	89%
Hyalella mortality	**0.09**	**22%**	**11%**	**78%**	**89%**	**37%**	**98%**	**88%**
	0.60	56%	1%	44%	99%	80%	96%	95%
Hyalella pooled	**0.28**	**36%**	**35%**	**64%**	**65%**	**41%**	**83%**	**65%**
	0.60	88%	1%	13%	99%	89%	75%	76%

[a] L1, L2, and L3 are 90%, 80%, and 70% differences, respectively.
Non-shaded rows indicate fixed PrMax threshold of 0.6; shaded rows with **bold** text indicate error-optimized PrMax threshold.

Location of errors within the Study Area

The error rates associated with Effects Level 1 are fairly high (one or both greater than 30%; Table 5-13). Similarly, the error rates for *Hyalella* pooled (driven by the contribution of the *Hyalella* growth endpoint) are also fairly high at all effects levels (Table 5-13). Error rates using the optimal threshold (shaded rows in Table 5-13) for *Chironomus* pooled and *Hyalella* mortality for Effects Levels 2 and 3 are better (< 26%). However, with predicted hit reliabilities less than 40%, this means that 60% of the stations predicted to be toxic are not toxic. The hit reliability could be improved by increasing the PrMax threshold, although this comes at the cost of increasing false negatives above 50%.

Figures 5-10 and 5-11 show the locations of errors associated with the LRM for *Chironomus* pooled and *Hyalella* mortality endpoints at Level 2 and Level 3, respectively. On these figures, toxicity was predicted if the calculated PrMax value for a station exceeded the optimal threshold (screened rows in Table 5-13). These figures illustrate false positives (stations without observed toxicity and PrMax values above the threshold) and false negatives (stations with observed toxicity and PrMax values below the threshold). At both of these levels, false negatives are rare and fairly randomly spread throughout the area. False positives are also scattered throughout the area, but there are a few clusters of false positives that are worth noting:

- Along the shore just southwest of RM 9
- In Swan Island Lagoon
- Between RM 6 and 7.5 on the west bank of the river
- On the north shore, north of Cathedral Park
- On the east side of the river just south of RM 2

Stations that exceed the PrMax for Levels 2 and 3 are presented in Figures 5-12 and 5-13, respectively. It should be noted that these figures show all surface sediment stations.

Chemical drivers

The chemicals associated with toxicity through the LRM were identified as those chemicals that had a high predicted probability of toxicity (max_p value > 0.60) at stations that were actually toxic. The chemicals are listed in Table 5-14, from the most important (predicting the most hits accurately) to least important within each endpoint. The list varies by endpoint somewhat, although there are some similarities. For example, diesel-range hydrocarbons and other organics are high on the list for pooled *Chironomus* and *Hyalella* mortality at all levels. Percent fines and the chemical endpoints correlated with percent fines (e.g., ammonia, aluminum, selenium) are high on the list for *Hyalella* pooled at Levels 1 and 2. The list of chemicals predicting Level 3 *Hyalella* pooled response is more similar to the list for *Chironomus* and *Hyalella* mortality.

Table 5-14. Chemicals responsible for accurate predictions of toxicity

	Chemicals			
	***Chironomus* Pooled**	***Hyalella* Mortality**	***Hyalella* Pooled**	
Level 1	diesel-range hydrocarbons	diesel-range hydrocarbons	percent fines	lead
	sulfide	sulfide	ammonia	silver
	dibutylphthalate	naphthalene	copper	beta-hexachlorocyclohexane
	4-methylphenol	residual-range hydrocarbons	sulfide	2-methylnaphthalene
	total DDE	total chlordane	selenium	delta-hexachlorocyclohexane
	lead	total DDE	aluminum	dibenzofuran
	mercury	total DDT	mercury	alpha-hexachlorocyclohexane
	carbazole		total chlordane	
	total chlordane		tributyltin	
	phenol		arsenic	
	dibenzofuran		pentachlorophenol	
	zinc		diesel-range hydrocarbons	
	tributyltin		naphthalene	
	selenium		phenol	
	copper		antimony	

Table 5-14. Chemicals responsible for accurate predictions of toxicity

	Chemicals			
	Chironomus **Pooled**	***Hyalella*** **Mortality**	***Hyalella*** **Pooled**	
Level 2	diesel-range hydrocarbons	beta-hexachlorocyclohexane	percent fines	cadmium
	sulfide	diesel-range hydrocarbons	selenium	naphthalene
	dibutylphthalate	naphthalene	aluminum	delta-hexachlorocyclohexane
	total DDE	sulfide	ammonia	total dioxins/furans
	dibenzofuran	total chlordane	beta-hexachlorocyclohexane	lead
	lead	total DDE	silver	nickel
	carbazole	total DDT	tributyltin	total chlordane
	total chlordane		phenol	zinc
	antimony		sulfide	residual-range hydrocarbons
	mercury		dibutylphthalate	total DDE
	4-methylphenol		antimony	total DDT
			copper	TEQ mammal (RL = 0.5 RL)
Level 3	diesel-range hydrocarbons	naphthalene	diesel-range hydrocarbons	total DDE
	sulfide	total DDE	aluminum	sulfide
	total DDE	diesel-range hydrocarbons	alpha-hexachlorocyclohexane[a]	arsenic
	residual-range hydrocarbons	residual-range hydrocarbons	naphthalene	total chlordane
	carbazole	total chlordane	phenol	tributyltin
	total chlordane	total DDT	copper	antimony
	total DDT	sulfide	silver	lead
	dibenzofuran	beta-hexachlorocyclohexane	nickel	beta-hexachlorocyclohexane
	mercury		zinc	mercury
			total DDT	selenium
				TEQ mammal (RL = 0.5 RL)[a]

[a] Low confidence in this model (see Appendix E).
RL – reporting limit

Influence of grain size

The strength of the relationship between percent fines and toxicity can be observed in the individual regression models (Appendix E). An effect of grain size on toxicity is seen only for *Hyalella* pooled at Levels 2 and 3. This correlation between the *Hyalella* pooled and percent fines is indicated by the presence of percent fines as a chemical driver.

5.4 DISCUSSION OF CHEMICAL DRIVERS

Both the LRM and the FPM found that the chemicals associated with toxicity vary by bioassay endpoint. While there were small differences between the models in terms of the exact analytes identified, the similarities were much greater. Minor differences are expected when chemicals covary in a data set, inasmuch as the specific analytes that each model selects may actually represent a larger group of analytes. This is particularly

noticeable among metals at this and other sites; the covariance also occurs among many organic chemical classes but is typically dealt with by summing these classes. The two models used different mathematical approaches, which, among other things, worked best with different approaches to pooling the endpoints. Therefore, some differences are not unexpected. However, the major drivers are similar, as discussed below.

The primary LRM results are based on the *Chironomus* pooled and *Hyalella* mortality endpoints, while the FPM uses *Chironomus* growth and mortality and *Hyalella* mortality. Together, they identified bulk hydrocarbons, PAHs, ammonia, sulfides, mercury, DDTs, chlordanes, di-n-butyl phthalate, and hexachlorocyclohexane as the primary chemical drivers for the Study Area. Lead was also identified by the LRM, whereas cadmium, silver, and PCBs were identified by the FPM. As noted above, it is likely that these metals covary with each other and/or with mercury to some extent. The FPM's somewhat greater reliability may also derive in part from incorporating these additional analytes into the model (e.g., PCBs).

Similar results were seen for the *Hyalella* growth and pooled endpoints in the FPM and the *Hyalella* pooled endpoint in the LRM. Although these endpoints are not recommended for use, in part because both models identified conventionals (fines, ammonia, and sulfides) as their primary chemical drivers. In addition, both models indicated that *Hyalella* growth is weakly responsive to a few additional metals, though again, not always the same ones.

Both models identified ammonia and sulfides as analytes associated with toxicity in this data set. Ammonia and sulfides are common confounding factors in bioassays (ASTM 2003) and can sometimes be high enough to cause toxicity in bulk sediments, even when their levels in overlying water are below bioassay QA/QC criteria. Ammonia and sulfides in sediments are formed as a result of bacterial action on decaying organic matter, which is a natural process. The source of the organic matter may be natural, particularly in backwater fine-grained areas, or it may be anthropogenic. In addition, both ammonia and sulfides can be present in some anthropogenic source materials as well as naturally produced in sediments. Detailed evaluation of the pattern of ammonia and sulfides concentrations with respect to both natural features and anthropogenic sources will be needed as part of the ERA to evaluate the nature of and appropriate response to this observed effect.

6.0 SUMMARY AND CONCLUSIONS

The findings of the study to identify a predictive model to be used in assessing risk to benthic invertebrates in the ERA for the Portland Harbor Superfund Site are presented below.

6.1 METHODS NOT RETAINED FOR USE

This section summarizes methods, endpoints, and effects levels that were evaluated for use but are not proposed as part of the final model. The rationale for each recommendation is presented below.

- **Existing SQV Sets and Site-specific AETs.** Five existing SQV sets used in North America and two quotient methods were evaluated to determine whether they would be reliable in predicting benthic toxicity in Portland Harbor. Most existing SQV sets at Levels 1 and 2 classified nearly all stations in the Harbor as hits, even though the majority of the bioassays showed no effects. Error rates were more evenly balanced between false negatives and false positives at Level 3, but both types of errors were well above 20%. Two likely reasons for these errors exist. First, most of these methods use relatively simplistic mathematical models compared to the FPM or the LRM. Second, the existing SQVs were generally based on acute toxicity data with a limited suite of biological endpoints, often incorporating data of varying quality from many different regions. Both the FPM and the LRM achieved substantially better performance than the existing SQVs; therefore, the existing SQV sets were not retained for use.

  Site-specific AETs for Portland Harbor were also calculated and evaluated. While most of the other reliability parameters were within acceptable ranges, a significant concern was that the false negatives were very high, ranging from 60 to 90%. Past evaluations conducted for the Washington Department of Ecology (Avocet and SAIC 2002), Port of Portland and ODEQ (unpublished) have also shown that freshwater AETs are frequently less reliable and far less conservative than marine AETs. The reasons for this are unknown, but it may have to do with the more variable bioavailability of metals in freshwater environments, leading to greater overlap between their hit and no-hit distributions. The bioavailability and toxicity of other chemicals, such as ammonia and ionic organic chemicals, may also be more variable in freshwater than in marine environments, where salinity and pH is buffered. For these reasons, the site-specific AETs are not proposed for use.

- ***Hyalella* Growth Endpoint.** In developing the model, it became clear that the *Hyalella* growth endpoint was responding differently than the

other endpoints from a variety of standpoints, which raised some concerns.

- **Lack of Correlation to Chemicals of Concern.** All three of the other endpoints responded strongly to PAHs and petroleum, various metals, and several organic chemicals and chemical classes of concern, along with ammonia and sulfides. However, the *Hyalella* growth endpoint was correlated most strongly with percent fines and ammonia and had only weak correlations with a few metals. This pattern strongly affected the pooled endpoint and also made the pooled endpoint less sensitive to chemicals of concern.
- **Poor reliability.** The *Hyalella* growth endpoint had substantially lower reliability than did the other three endpoints in all three site-specific methods evaluated – the FPM, LRM, and site-specific AETs. This was the only endpoint that was not capable of reliably predicting toxicity in Portland Harbor sediments at Levels 2 and 3. Pooling this endpoint with *Hyalella* mortality, which was otherwise quite reliable, also reduced the reliability of the pooled endpoint below acceptable levels.
- **Effect of Percent Fines.** *Hyalella* growth (and the associated pooled endpoint) appears to be the only endpoint affected by grain size, with effects beginning at approximately 60% fines. As discussed in Section 5.2, AETs for percent fines were also calculated. For all other toxicity endpoints (i.e., *Hyalella* mortality and *Chironomus* mortality and growth), the AET was 100%; but for this endpoint, the AET was approximately 80% fines. The results for both the FPM and the AET methods indicate some level of adverse effects of high fines on the growth endpoint. *Hyalella* growth was more strongly associated with fines than with any other analytical parameter, with the possible exception of ammonia. At the same time, percent fines was not significantly correlated with toxic COPCs at the Study Area. Neither *Hyalella* nor *Chironomus* are currently thought to be significantly influenced by percent fines (Ankley et al. 1994; Ingersoll et al. 1996). However, most of the testing with *Hyalella* has been with the mortality endpoint. The use of the growth endpoint (with the associated longer exposure time) has been a relatively recent addition to toxicity testing. There is not much of a track record with this test in the region to date, and it seems appropriate to raise the possibility that there is an effect of sediment with very high percent fines on growth in the long-term test that has previously gone unrecognized. Certainly, there are precedents for high- and low-percent fines effects on other amphipods, both freshwater and marine, in commonly used toxicity tests. The poor reliability of the *Hyalella* test in predicting the toxicity associated with chemical concentrations may be because of the confounding effects of grain

size on the results, since grain size is not well-correlated with chemical concentrations in this data set.

- **Correlation among Endpoints.** The relationship among toxicological endpoints is such that there is very little correlation between *Hyalella* growth and mortality (Spearman $\rho = -0.09$, $p = 0.19$), whereas *Chironomid* growth and mortality are strongly correlated (Spearman $\rho = 0.37$, $p = 0$; Figure 4-3). It appears that the *Hyalella* growth endpoint has a distinctly different response to sediment characteristics than do the other three endpoints: there is a lack of correlation between *Hyalella* growth and the other three toxicological endpoints, and there is a lack of correlation between the *Hyalella* growth or pooled endpoints and COPCs. This is inconsistent with the correlation observed between the other three toxicological endpoints and the COPCs.

  In summary, the *Hyalella* growth endpoint largely does not respond to COPCs at the Study Area and has no relation to any other endpoint in its patterns of response. Its reliability is poor at all effects levels and greatly reduces the reliability of the pooled endpoint. The *Hyalella* growth endpoint seems to be responding primarily to percent fines and ammonia. For these reasons, it is recommended that this endpoint, as well as the pooled *Hyalella* endpoint, not be used in developing a predictive model or SQVs for Portland Harbor. An effective model can be built using the other three individual endpoints or by using the *Chironomus* pooled endpoint and the *Hyalella* mortality endpoint.

- **Level 1 Biological Effects Level.** The reliability of nearly all the endpoints at Level 1 is reduced as compared to Levels 2 and 3. This is likely due to the very small difference (10%) from control used to define the Level 1 endpoints. This level of difference is likely within natural and laboratory variability in many cases and is smaller than the MDD reported for many of these endpoints in round robin tests conducted for the American Society for Testing and Materials (ASTM) protocols (ASTM 2003). Appendix A, Table A-2, presents the numbers of statistically indeterminate stations, and there are significantly more indeterminate results at Level 1 than at the other two effects levels. Because of these natural variability and statistical issues, it is unlikely that any SQV set could perform with high reliability in predicting these very small variations in effects.

  Effects levels this low are not known to have been adopted by any regulatory program for the protection of benthic organisms, inasmuch as it is not clear that these levels can be reliably measured for most endpoints or that population-level effects actually occur due to small variations that are within natural variability. In a regional context, both Washington State and British Columbia have adopted SQVs with lower levels set at

approximately 20% effects (equivalent to Level 2 in this study) and upper levels set between 30 and 50% effects (at or above Level 3 in this study).

Therefore, it is recommended that Level 1 not be used to set SQVs for Portland Harbor because it is relatively unreliable in accurately predicting effects and well below the cleanup levels set at other regional Superfund sites. Levels 2 and 3 are as or more conservative than levels used in state programs, federal Superfund programs, and regional dredging programs and have good reliability in predicting both acute and chronic toxicity in sediments.

6.2 LOGISTIC REGRESSION MODEL

The overall utility of this method for predicting toxicity from chemistry is fairly limited as indicated by the high error rates and poor reliability outcomes (see Section 5.3). The exploratory analysis indicates that there is very little relationship between chemical concentrations and toxicity. The errors (false positives and false negatives) associated with using a single PrMax threshold to define a clear line between stations predicted as toxic or non-toxic cannot be simultaneously maintained at a reasonable level. The results from this model may be useful to illustrate the spatial distribution of toxicological risk as a result of combined chemical concentrations. As shown in Figure 6-1, areas with the highest PrMax values may be at potentially higher risk, while areas with the lowest PrMax values may be at potentially lower risk of toxicity. The areas with higher PrMax values generally confirm the results of the FPM SQVs (see Section 6.3).

The following conclusions can be drawn from the development of the LRM models:

- **Chemicals associated with toxicity vary by endpoint.** The chemicals that were most associated with toxicity with the LRM were identified as those that set a max_p value > 0.60 for toxic stations (Table 5-14). The list varied somewhat by endpoint.
  - For the *Chironomus* pooled endpoint, the strongest relationships exist with diesel-range hydrocarbons, PAH-like compounds (i.e., carbazole and dibenzofuran), sulfide, certain metals (i.e., lead and mercury), and specific organics (DDE, chlordane, and di-n-butyl phthalate).
  - For the *Hyalella* mortality endpoint, the strongest relationships exist with diesel- and residual-range hydrocarbons (i.e., bulk hydrocarbons), PAHs (e.g., naphthalene), sulfide, and certain other organics (hexachlorocyclohexane, chlordane, DDE, and total DDTs).
  - The *Hyalella* pooled endpoint had the strongest relationships between toxicity and percent fines, ammonia, sulfide and individual metals (i.e., aluminum, selenium, copper, and mercury); other

chemicals that were also associated with toxicity included diesel-range hydrocarbons and naphthalene, other organics (hexachlorocyclohexane, di-n-butyl phthalate, chlordane, and total DDTs), phenols (e.g., phenol and pentachlorophenol), other metals (e.g., antimony, arsenic, cadmium, lead, nickel, silver), and tributyltin (as ion).

- **Individual LRMs were developed for both individual and pooled endpoints.** LRMs were developed for the pooled *Chironomus*, pooled *Hyalella*, as well as the *Hyalella* mortality endpoint, each at three different effects levels. Individual chemical models were developed for up to 67 individual chemicals for each biological endpoint. For each biological endpoint, a multi-chemical model was developed to predict the probability of toxicity based on the suite of chemical mixtures at a station.
- **Effects Level 1 exhibits the highest error rates and lowest reliability for *Chironomus* pooled and *Hyalella* mortality.** Reliability results for the LRM model were similar to those for the FPM model, with Level 1 models exhibiting much higher up to 15% error rates and/or up to 9% lower reliabilities than Levels 2 and 3.
- **Reliability of the LRM was high for two out of three endpoints.** The LRM showed good performance in predicting toxicity for *Hyalella* mortality at Effects Levels 2 and 3. The performance for the pooled *Chironomus* endpoint was also fairly good (error rates < 26%) at Levels 2 and 3. Performance for the pooled *Hyalella* endpoint was poor.

6.3 FLOATING PERCENTILE MODEL

Because it has the greatest reliability in predicting benthic toxicity, the FPM is recommended for use in developing site-specific SQVs for Portland Harbor. The following key results and conclusions were identified during the development of this model and the associated SQVs:

- **There is a limited set of chemicals associated with toxicity.** A total of 38 chemicals or chemical classes had more than 30 detections in the data set and were evaluated for inclusion in the FPM. Of these, 20 were found to have a significant relationship with at least one measure of toxicity in the data set, as determined by an ANOVA comparison of their hit and no-hit distributions. Of these 20, between 7 and 14 chemicals were significant for any one individual biological endpoint.
- **Sensitivity to individual chemicals varies by endpoint.** The chemicals that showed a relationship to toxicity varied by endpoint. The *Chironomus* growth, *Chironomus* mortality, and *Hyalella* mortality endpoints were sensitive to similar chemicals, while the *Hyalella* growth

endpoint showed a very different relationship. For most endpoints, the strongest relationship with toxicity exists for bulk hydrocarbons, PAHs, ammonia and sulfides, certain metals (e.g., cadmium, mercury, silver), and certain other organics (hexachlorocyclohexane, PCBs, DDTs, chlordane, di-n-butyl phthalate). The *Hyalella* growth endpoint has strong relationships only with percent fines and ammonia and has weak relationships with certain metals (i.e., copper, arsenic, nickel, zinc).

- **FPM SQVs were developed for both individual and pooled endpoints.** Chemical SQVs were developed for each of the four endpoints using the chemicals associated with each specific endpoint. In addition, pooled models were developed for the two *Chironomus* endpoints and the two *Hyalella* endpoints. SQVs were developed for all three effects levels (Levels 1, 2, and 3).
- **Reliability of the FPM model was high for three out of four endpoints.** The FPM showed good performance in predicting toxicity for three out of the four biological endpoints (i.e., *Chironomus* mortality and growth and *Hyalella* mortality). The *Hyalella* growth endpoint showed poor performance, as might be expected since adverse effects in this endpoint appear to be primarily related to conventional parameters (percent fines and ammonia) rather than to toxic COPCs. An approach that uses the lowest of the SQVs for the other three endpoints is recommended as an indication of potential risk to the benthic community.
- **Effects Levels 2 and 3 can be reliably predicted and are recommended for use in Portland Harbor.** Reliability of the FPM was greater at Effects Levels 2 and 3 than at Effects Level 1. Level 1 had some stations that were statistically indeterminate and may be too low an effects level to predict reliably. Levels 2 and 3 are conceptually consistent with levels that have been adopted for cleanup within EPA Region 10 and in other states and provinces in North America.
- **Results of the model are geographically consistent with known sources.** Figure 6-2 identifies stations that exceed the FPM pooled SQVs for the three recommended endpoints at Levels 2 and 3. Clusters of exceedances clearly identify specific areas of predicted benthic toxicity within Portland Harbor along both banks of the river that are related to known upland sites and sources. The results of the model correspond well with both measured toxicity and the conceptual site model.
- **There are a few areas where additional toxicity testing may be warranted.** In a few areas, mapping of errors indicates that the model may over-predict toxicity, most likely due to higher concentrations of chemicals in matrices that are less bioavailable, such as paint chips or weathered petroleum. In these areas, biological testing should be an

option during the remedial design process to confirm any predictions using the SQVs.

6.4 PROPOSED SEDIMENT QUALITY VALUES

Proposed Level 2 and Level 3 SQVs are presented in Table 6-1. These SQVs represent the lowest of the SQVs for the three recommended endpoints at each level of effects.

Table 6-1. Proposed Effects Level 2 and Effects Level 3 SQVs

ANALYTE	UNITS	LEVEL 2 SQVS	LEVEL 3 SQVS
Ammonia	mg/kg	170	280
Sulfides	mg/kg	32	415
Arsenic	mg/kg	24	34
Cadmium	mg/kg	2.6	2.6
Copper	mg/kg	562	562
Mercury	mg/kg	0.63	0.63
Silver	mg/kg	32	415
beta-Hexachlorocyclohexane	µg/kg	9.6	9.6
Dieldrin	µg/kg	21.5	21.5
Diesel-range hydrocarbons	µg/kg	340,000	340,000
Di-n-butyl phthalate	µg/kg	90	90
Residual-range hydrocarbons	µg/kg	2,700,000	4,500,000
Total DDTs	µg/kg	1,000	1,000
Total PAHs	µg/kg	1,270,000	1,270,000
Total PCBs	µg/kg	1,400	1,450

Chemicals were not included in the list of SQVs if the value assigned by the FPM was the highest concentration in the data set (equivalent to a "greater than" AET). In other words, the actual toxicity threshold is unknown but is above the concentration distribution in this data set. These chemicals include percent fines, antimony, chromium, lead, nickel, zinc, methoxychlor, total chlordane, delta-hexachlorocyclohexane, hexachlorobenzene, 4-methylphenol, pentachlorophenol, bis(2-ethylhexyl phthalate), butylbenzyl phthalate, monobutyltin, dibutyltin, tributyltin, tetrabutyltin, and total dioxins. These chemicals are not likely to be important in identifying benthic toxicity in this data set at Levels 2 and 3.

The FPM and the LRM identify a relatively limited suite of metals and organics, as well as ammonia and sulfides, associated with toxicity. Each of these may be representing other chemicals that are co-located and/or of lower toxicity than the ones included in the SQV set. Together, the chemicals identified in Table 6-1 are reliable in predicting adverse effects to benthic communities in Portland Harbor.

An important point to note is the performance of bulk petroleum measures (diesel-range hydrocarbons and residual-range hydrocarbons) as compared to individual and total

PAHs. Bulk petroleum measures were more strongly correlated with toxicity than total PAHs, even though PAHs were measured at all stations, and bulk petroleum was measured at only a subset of stations. Although the SQVs for PAHs may appear high, they are consistent with those derived from other West Coast data sets (e.g., San Francisco Harbor (Germano & Associates 2004), Los Angeles Harbor (unpublished)) using the FPM and the LRM, indicating that PAHs alone are not large contributors of toxicity to benthic organisms. PAHs are only a small subset of the suite of narcotic chemicals present in sediments and in petroleum, all of which may affect benthic organisms through similar toxicological pathways (McCarty 1991; McCarty and Mackay 1993; McCarty et al. 1992). The bulk measures of petroleum appear to better capture and correlate with that toxicity, as is apparent from the SQVs calculated for these measures.

The FPM often identifies similar values for different effects levels, as can be seen in Table 6-1 (this is also true of AETs). Some chemicals, such as ammonia, arsenic, and residual-range hydrocarbons, have different SQVs at Level 2 and Level 3. Other chemicals, such as copper, diesel-range hydrocarbons, and DDTs, have the same SQV at both levels. Although at first this may appear unusual, it reflects the fact that the concentration-toxicity curve for these chemicals is apparently steep in Portland Harbor. At the level at which the effects associated with these chemicals can be reliably seen, the effect is clear enough that it exceeds both Level 2 and Level 3.

A review of the bioassay results indicates that many of the same stations exceed both Level 2 and Level 3, which results in the pattern of site-specific SQVs observed in this analysis. From a practical standpoint, this creates a relatively clear distinction between areas that are not likely to experience effects and areas in which the benthic community may be at greater risk, without a large "grey zone" in between (see Figure 6-3 for a comparison of Levels 2 and 3).

7.0 RECOMMENDATIONS

The model development and analysis presented in this report demonstrates that a predictive benthic toxicity model can be developed for use in the Portland Harbor ERA. Site-specific SQVs with acceptable overall reliability that were able to minimize both false positive and false negative errors were developed. The range of biological effects levels are consistent with those used in other regulatory programs and will be useful in identifying risk of biologically meaningful adverse effects in the Study Area. While both the FPM and the LRM initially showed promise in predicting Portland Harbor-specific toxicity based on surface sediment concentrations, the analysis presented in this report indicates that the FPM would better meet the needs of the RI/FS being conducted for the Portland Harbor Superfund Site.

As presented in Section 1.0, the predictive model will be used for two primary purposes, namely to identify:

- SQVs that reliably predict benthic toxicity in the Study Area
- Areas within Portland Harbor where sediment chemical concentrations pose a risk to benthic invertebrates

The FPM is a useful tool for identifying surface sediments that may be potentially toxic to benthic invertebrates. Based on the analysis of predictive reliability of the three proposed effects levels, Effects Levels 2 and 3 appear to best fit an operating definition for assessing risks to the benthic community and give results consistent with the geographic distribution of COPCs and known sources.

The Effects Level 2 definition is similar to the operational definition recommended by ASTM for determining when a toxicity test response is significantly different from reference samples for freshwater toxicity tests. It is also similar to the lower-tier response levels used in regulatory decision-making by various jurisdictions (e.g., analogous to the SQS in the State of Washington Sediment Management Standards and the effects level used by British Columbia for sensitive aquatic areas). Effects Level 3 is similar to the CSL in the Washington State Sediment Management Standards and the effects level used by British Columbia for urban harbors. Effects levels within the Level 2 and Level 3 range have been applied by EPA at a number of Superfund sites in the Pacific Northwest, such as Commencement Bay, the Duwamish River, Eagle Harbor, and Ketchikan Pulp Co. The Level 2 and Level 3 SQVs listed in Table 6-1 were used to develop Figure 6-3. From this figure, it is evident that the application of the FPM at either Levels 2 or 3 identifies distinct areas of potential risk to benthic communities based on clustered locations with either observed or predicted toxicity (hit locations). Consistent with how these levels have been used in other jurisdictions, Level 3 might provide more compelling evidence of benthic toxicity, while Level 2 could be used in conjunction with other LOEs to establish areas of concern.

The FPM SQVs can be used to identify and map sediments with predicted benthic toxicity within the Study Area. This approach can be easily applied at varying scales to support an analysis of potential impacts on the benthic community. It can also be used to gain a site-wide perspective (see Figure 6-3 for an example of this) or can be used to evaluate much smaller scales, including the potential for toxicity on a point-by-point basis or the identification of AOPCs. The purpose of this report is not to identify specific AOPCs related to benthic toxicity; however, these will be identified in the ERA based on a variety of factors, including:

- Exceedance of bioassay toxicity thresholds
- Exceedance of site-specific SQVs (at stations without bioassay data)
- Grouping of individual stations with exceedances into areas of benthic toxicity
- Information on chemical similarity among groups of stations and known sources and transport pathways to sediments

The predictive model can also be used in post-Record of Decision (ROD) remedial design decisions in areas in which direct toxicity to benthic organisms is an important consideration for risk reduction. Either the site-specific SQVs or the associated bioassay effects levels can be applied to any additional surface sediment data collected during the design phase to aid in further defining remedial boundaries. Bioassay testing would be particularly appropriate in areas where the mapping of errors indicates that false positives may be likely. If toxicity tests are conducted, then the Effects Level 2 or 3 hit definitions would be used to determine if the resulting test response data represents a toxic sample.

Finally, either direct bioassay testing or the site-specific SQVs can be used in post-remediation monitoring to ensure that the selected remedy continues to be protective of the benthic community.

8.0 REFERENCES

Ankley GT, Benoit DA, Balouth JC, Reynoldson TB, Day KE, Hoke RA. 1994. Evaluation of potential confounding factors in sediment toxicity tests with three freshwater benthic invertebrates. Environ Toxicol Chem 13:627-635.

ASTM. 2003. Standard method for measuring the toxicity of sediment associated contaminants with freshwater invertebrates. E 1706–00. American Society for Testing and Materials, West Conshohocken, PA.

Avocet. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: Development and recommendation of SQVs for freshwater sediments in Washington State. Publication No. 03-09-088. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA.

Avocet, SAIC. 2002. Development of freshwater sediment quality values for use in Washington State. Final report. Publication No. 02-09-050. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA and Science Applications International Corporation (SAIC), Bothell, WA.

CCME. 2002. Canadian sediment quality guidelines for the protection of aquatic life: summary tables. Canadian Environmental Quality Guidelines, 1999, updated 2001, updated 2002 [online]. Canadian Council of Ministers of the Environment. Updated 2002. [Cited March 2005]. Available from: http://www.ccme.ca/assets/pdf/e7_002.pdf.

Connell D, Markwell RD. 1992. Mechanism and prediction of nonspecific toxicity to fish using bioconcentration characteristics. Ecotoxicol Environ Saf 24:247-265.

Ecology. 2002. Development of freshwater sediment quality values for use in Washington State. Phase I task 6: final report. Publication no. 02-09-050. Toxics Cleanup Program, Washington Department of Ecology, Olympia, WA.

Ecology. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: development and recommendation of SQVs for freshwater sediments in Washington State. Publication no. 03-09-088. Toxics Cleanup Program, Washington Department of Ecology, Olympia, WA.

EPA. 2005a. Memo to Lower Willamette Group. Subject: benthic interpretive approach for Portland Harbor Ecological Risk Assessment. October 26, 2005. US Environmental Protection Agency, Portland, OR.

EPA. 2005b. Predicting toxicity to amphipods from sediment chemistry. EPA/600/R-04/030. National Center for Environmental Assessment, US Environmental Protection Agency, Washington, DC.

Field J, Norton S, MacDonald D, Severn C, Ingersoll C. 1999. Beyond thresholds: using logistic regression models to estimate the probability of toxicity from sediment chemistry. SETAC poster.

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn C, Smorong DE, Lindskoog RA. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ Toxicol Chem 9:1993-2005.

Germano & Associates. 2004. An evaluation of existing sediment screening guidelines for wetland creation/beneficial reuse of dredged material in the San Francisco Bay Area along with a proposed approach for alternative guideline development. Prepared for the California State Coastal Conservancy and the Port of Oakland. Germano & Associates, Inc., Bellevue, WA.

Ingersoll C, Brunson E, Dwyer J. 1996. Methods for assessing bioaccumulation of sediment-associated contaminants with freshwater invertebrates. USEPA National Sediment Bioaccumulation Conference.

Ingersoll CG, MacDonald DD, Wang N, Crane JL, Field LJ, Haverland PS, Kemble NE, Lindskoog RA, Severn C, Smorong DE. 2000. Prediction of sediment toxicity using consensus-based freshwater sediment quality guidelines. US Environmental Protection Agency, Chicago, IL.

Integral. 2005a. Portland Harbor remedial investigation/feasibility study round 2 surface and beach sediment field sampling report. Draft. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA.

Integral. 2005b. Portland Harbor remedial investigation/feasibility study round 2a sediment site characterization summary report. Draft. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA.

Integral, Windward, Kennedy/Jenks, Anchor, Groundwater Solutions. 2004. Portland Harbor remedial investigation/feasibility study programmatic work plan. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA; Windward Environmental LLC, Seattle, WA; Kennedy/Jenks Consultants, Portland, OR; Anchor Environmental, LLC, Seattle, WA; Groundwater Solutions, Inc., Portland, OR.

Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environ Toxicol Chem 17(4):714-727.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39:20-31.

McCarty LS. 1991. Toxicant body residues: implications for aquatic bioassays with some organic chemicals. In: Mayes MA, Barron MG, eds, Aquatic toxicology and risk assessment. Vol 14, ASTM STP 1124. American Society for Testing and Materials, Philadelphia, PA.

McCarty LS, Mackay D. 1993. Enhancing ecotoxicological modeling and assessment. Environ Sci Technol 27(9):1719-1728.

McCarty LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1992. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic organics. Environ Toxicol Chem 11:917-930.

Menard S. 2000. Coefficients of determination for multiple logistic regression analysis. Amer Statistician 54(1):17-24.

ODEQ. 1999. Portland Harbor sediment management plan. Appendix G - sediment assessment methodology. Oregon Department of Environmental Quality, Portland, OR.

Persaud D, Jaagumagi R, Hayton A. 1993. Guidelines for the protection and management of aquatic sediment quality in Ontario. ISBN 0-7729-9248-7. Ontario Ministry of Environment and Energy, Toronto, ON.

PSEP. 1988. Sediment quality values refinement: Volume I. Update and evaluation of Puget Sound AET. Prepared for Puget Sound Estuary Program (PSEP), US Environmental Protection Agency, Region 10. PTI Environmental Services, Inc., Bellevue, WA.

Veith GD, Call DJ, Brooke LT. 1983. Structure-toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. Can J Fish Aquat Sci 40:743-748.

Windward. 2005a. Portland Harbor RI/FS. Ecological risk assessment: Estimating risks to benthic organisms using sediment toxicity tests. WE-05-0003. Draft. Prepared for Lower Willamette Group. Windward Environmental LLC, Seattle, WA.

Windward. 2005b. Portland Harbor RI/FS. Round 2A data report: Sediment toxicity testing. WE-05-0001. Draft. Prepared for Lower Willamette Group. Windward Environmental LLC, Seattle, WA.

FIGURES

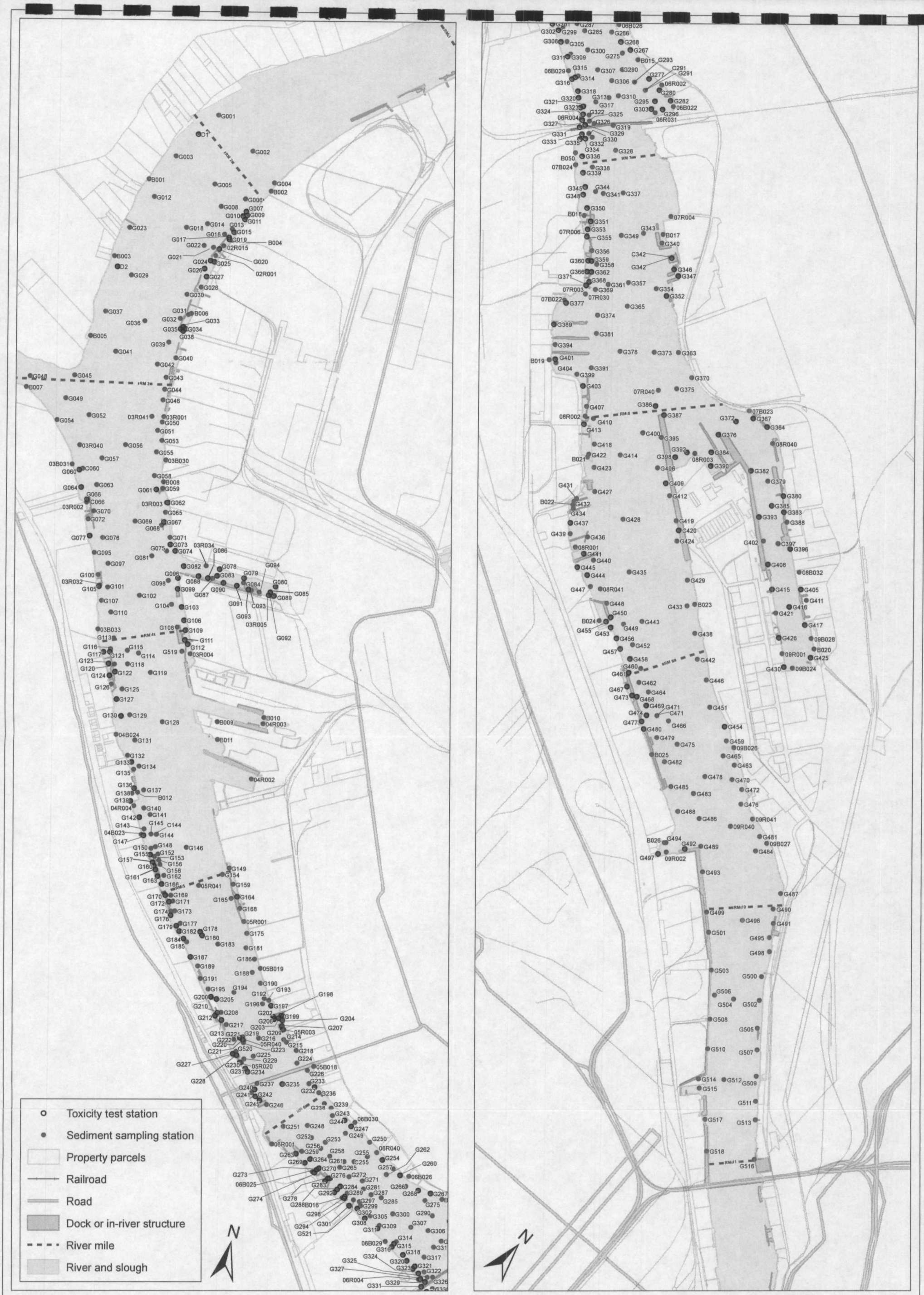


Figure 2-1. Portland Harbor Round 1 and 2 surface sediment and Round 2 toxicity test stations River Miles 2 to 11

Wind Ward environmental LLC

Prepared by AMW 02-03-06 Map 2169

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\Surface sediment and toxicity test stations

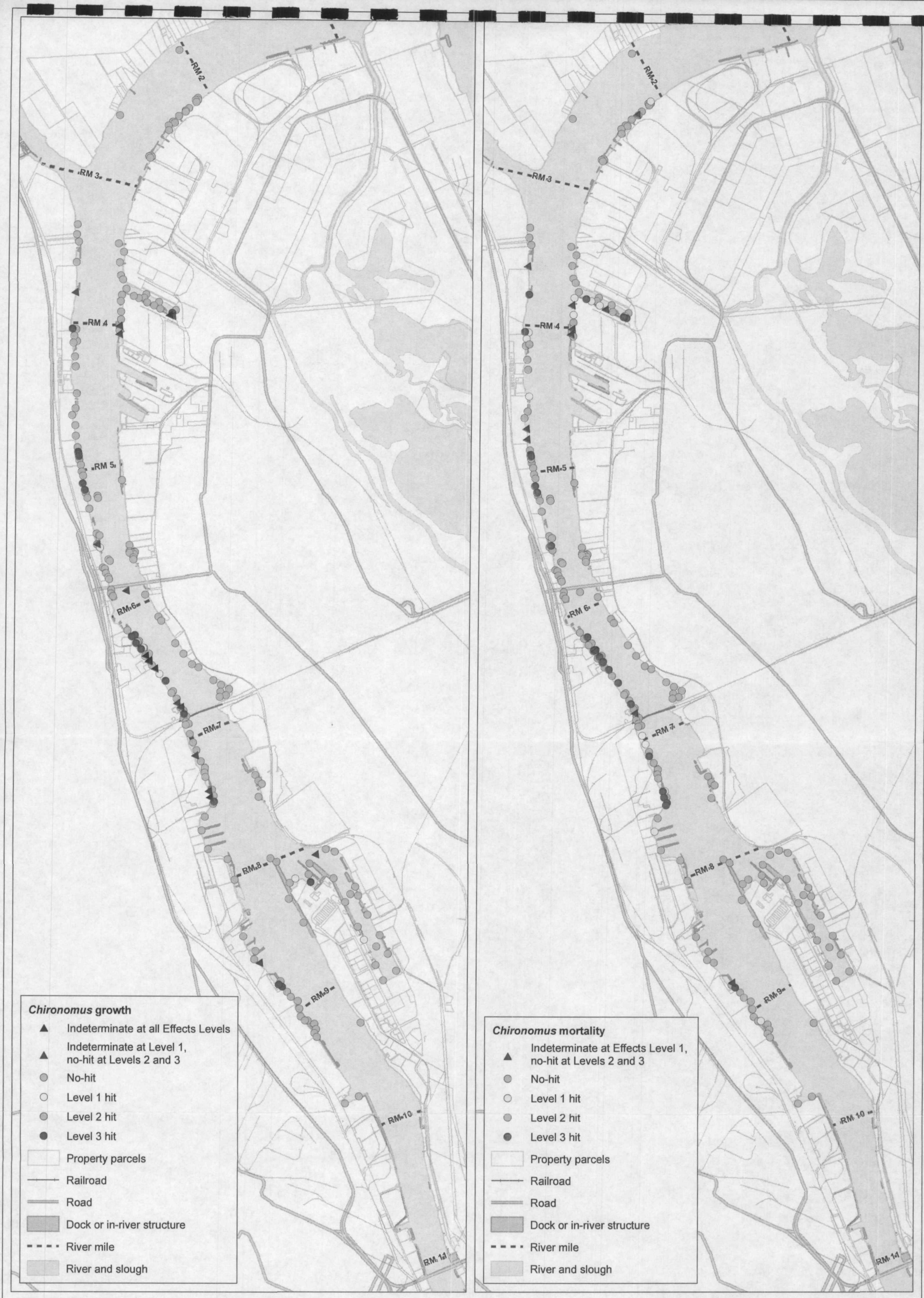


Figure 2-2. Toxicity test hit/no-hit and indeterminate stations for *Chironomus*

Wind Ward environmental LLC

Prepared by LSM 02-02-06 map 2166

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\

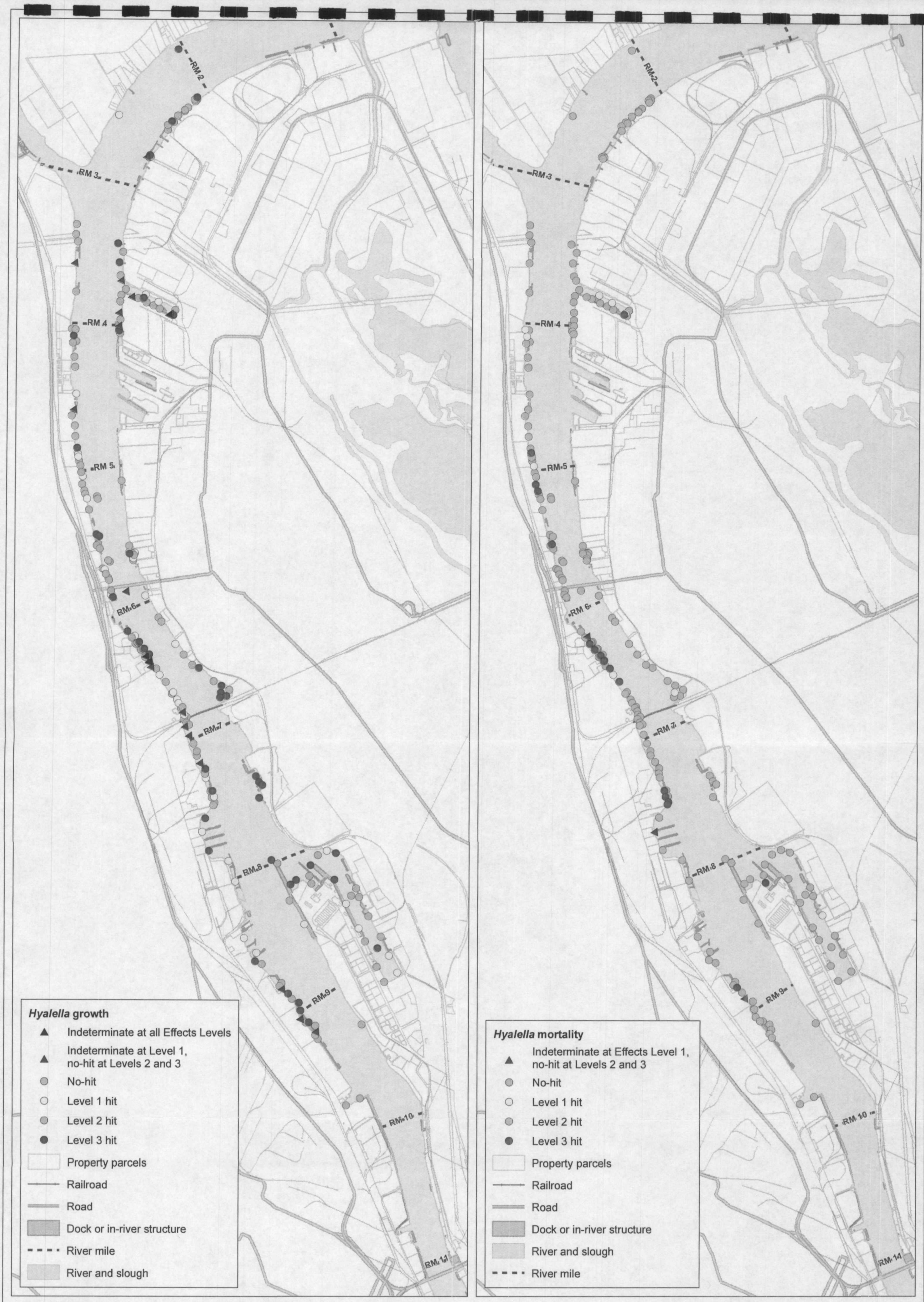
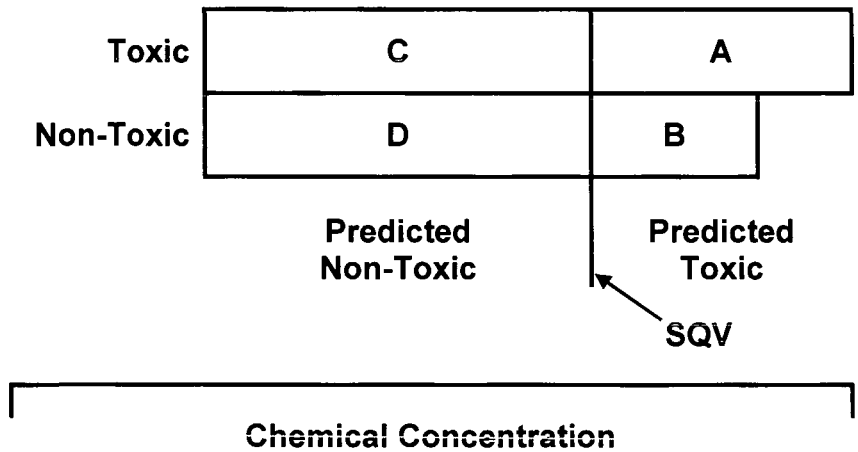


Figure 2-3. Toxicity test hit/no-hit and indeterminate stations for *Hyalella*

Prepared by LSM 02-02-06 map 2166

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\



False negatives = C/(A+C)
False positives = B/(B+D)
Sensitivity = A/(A+C)
Efficiency = D/(D+B)
Predicted hit reliability = A/(A+B)
Predicted no-hit reliability = D/(D+C)
Overall reliability = (D+A)/(A+B+C+D)

Figure 3-1. Calculation of reliability parameters

Figure 4-1. Correlations that have a significant Pearson's correlation coefficient

Group	Chemical	Fines (%)	TOC	Ammonia	Sulfide	Benzo(a)anthracene	Benzo(a)pyrene	Benzo(b)fluoranthene	Benzo(g,h,i)perylene	Benzo(k)fluoranthene	Chrysene	Dibenz(a,h)anthracene	Fluoranthene	Indeno(1,2,3-c,d)pyrene	Pyrene	Total HPAHs (calc'd)	2-Methylnaphthalene	Acenaphthene	Acenaphthylene	Anthracene	Fluorene	Naphthalene	Phenanthrene	Total LPAHs (calc'd)	Total PAHs (calc'd)	Carbazole	Dibenzofuran	Antimony	Arsenic	Cadmium	Chromium	Copper	Lead	Mercury	Nickel	Selenium	Silver	Zinc	Hexachlorobenzene	Total Dioxin Furans (calc'd)	Total PCBs Aroclors (calc'd)	Dibutyltin	Monobutyltin	Tetrabutyltin	Tributyltin	Diesel-Range Hydrocarbons	Aldrin	alpha-Hexachlorocyclohexane	beta-Hexachlorocyclohexane	delta-Hexachlorocyclohexane	Methoxychlor	Total Chlordane (calc'd)	Total DDTs (calc'd)	Total Endosulfan (calc'd)	4-Methylphenol	Pentachlorophenol	Phenol	Bis(2-ethylhexyl) phthalate	Butyl benzyl phthalate	Di-n-butyl phthalate
Conventional	Fines (%)																																																																		
	TOC																																																																					
	Ammonia																																																																				
	Sulfide																																																																			
HPAH	Benzo(a)anthracene						x	x	x	x	x	x	x	x	x			x	x	x	x		x	x	x	x	x																																										
	Benzo(a)pyrene					x		x	x	x	x	x	x	x	x				x	x	x		x	x	x																																											
	Benzo(b)fluoranthene					x	x		x	x	x	x	x	x	x				x	x	x		x	x	x	x																																									
	Benzo(g,h,i)perylene					x	x	x		x	x	x	x	x	x				x	x			x	x	x																																									
	Benzo(k)fluoranthene					x	x	x	x		x	x	x	x	x				x	x	x		x	x	x	x																																								
	Chrysene					x	x	x	x	x		x	x	x	x			x	x	x	x		x	x	x	x	x																																								
	Dibenz(a,h)anthracene					x	x	x	x	x	x		x	x	x				x	x	x		x	x	x	x	x																																							
	Fluoranthene					x	x	x	x	x	x	x		x	x			x	x	x	x		x	x	x	x	x																																							
	Indeno(1,2,3-c,d)pyrene					x	x	x	x	x	x	x	x		x				x	x			x	x	x																																									
	Pyrene					x	x	x	x	x	x	x	x	x				x	x	x	x		x	x	x	x	x																																							
HPAH sum	Total HPAHs (calc'd)					x	x	x	x	x	x	x	x	x	x			x	x	x	x		x	x	x	x	x																																							
LPAH	2-Methylnaphthalene																	x	x	x	x	x	x	x	x		x																																							
	Acenaphthene					x					x			x		x	x			x	x		x	x	x		x																																					
	Acenaphthylene					x	x	x	x	x	x	x	x	x	x	x	x			x			x	x	x																																							
	Anthracene					x	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x																																					
	Fluorene					x	x	x		x	x	x	x		x	x	x	x		x			x	x	x		x																																				
	Naphthalene																x							x																																							
	Phenanthrene					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x		x	x	x		x																																			
	Total LPAHs (calc'd)					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x																																				
PAH sum	Total PAHs (calc'd)					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x																																				
Pesticide	Carbazole					x		x		x	x	x	x		x	x					x					x																																				
Misc. organic	Dibenzofuran					x				x		x		x	x	x	x			x	x			x	x	x		x																																	
Metal	Aluminum																																																																			
	Antimony																																																																			
	Arsenic																																																																		
	Cadmium																																																																	
	Chromium																																																																	
	Copper																																																																	
	Lead																																																																	
	Mercury																																																																	
	Nickel																																																																	
	Selenium																																																																	
	Silver																																																																	
	Zinc																																																																	
Misc. organic	Hexachlorobenzene																																																														
	Total dioxins/furans (calc'd)																																																													
	Total PCBs Aroclors (calc'd)																																																													
Organotin	Dibutyltin																																																													
	Monobutyltin																																																													
	Tetrabutyltin																																																													
	Tributyltin																																																													
PAH sum	Diesel-Range Hydrocarbons																																																													
Pesticide	Aldrin																																																													
	alpha-Hexachlorocyclohexane																																																													
	beta-Hexachlorocyclohexane																																																													
	delta-Hexachlorocyclohexane																																																													
	Methoxychlor																																																													
	Total chlordane (calc'd)																																																													
	Total DDTs (calc'd)																																																													
	Total endosulfan (calc'd)																																																													
Phenol	4-Methylphenol																																																													
	Pentachlorophenol																																																													
	Phenol																																																														
Phthalate	Bis(2-ethylhexyl) phthalate																																																													
	Butyl benzyl phthalate																																																													
	Di-n-butyl phthalate																																																													

This figure summarizes the correlations that have a significant Pearson's correlation coefficient (minimum r: 0.9; maximum p value: 0.01).
Note: Only the upper triangle of the original matrix was filled; see intersections of rows and columns for important correlations with each chemical.

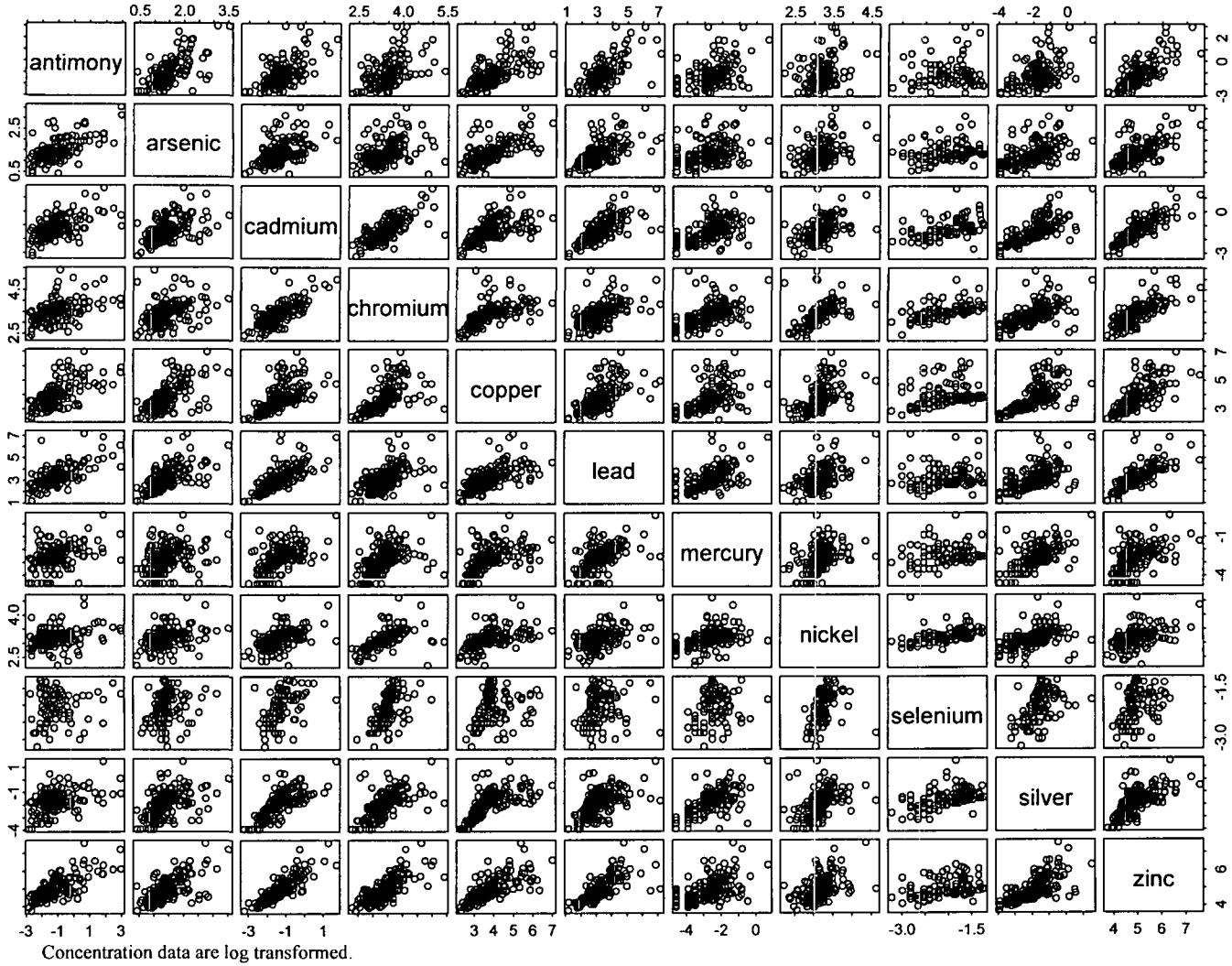
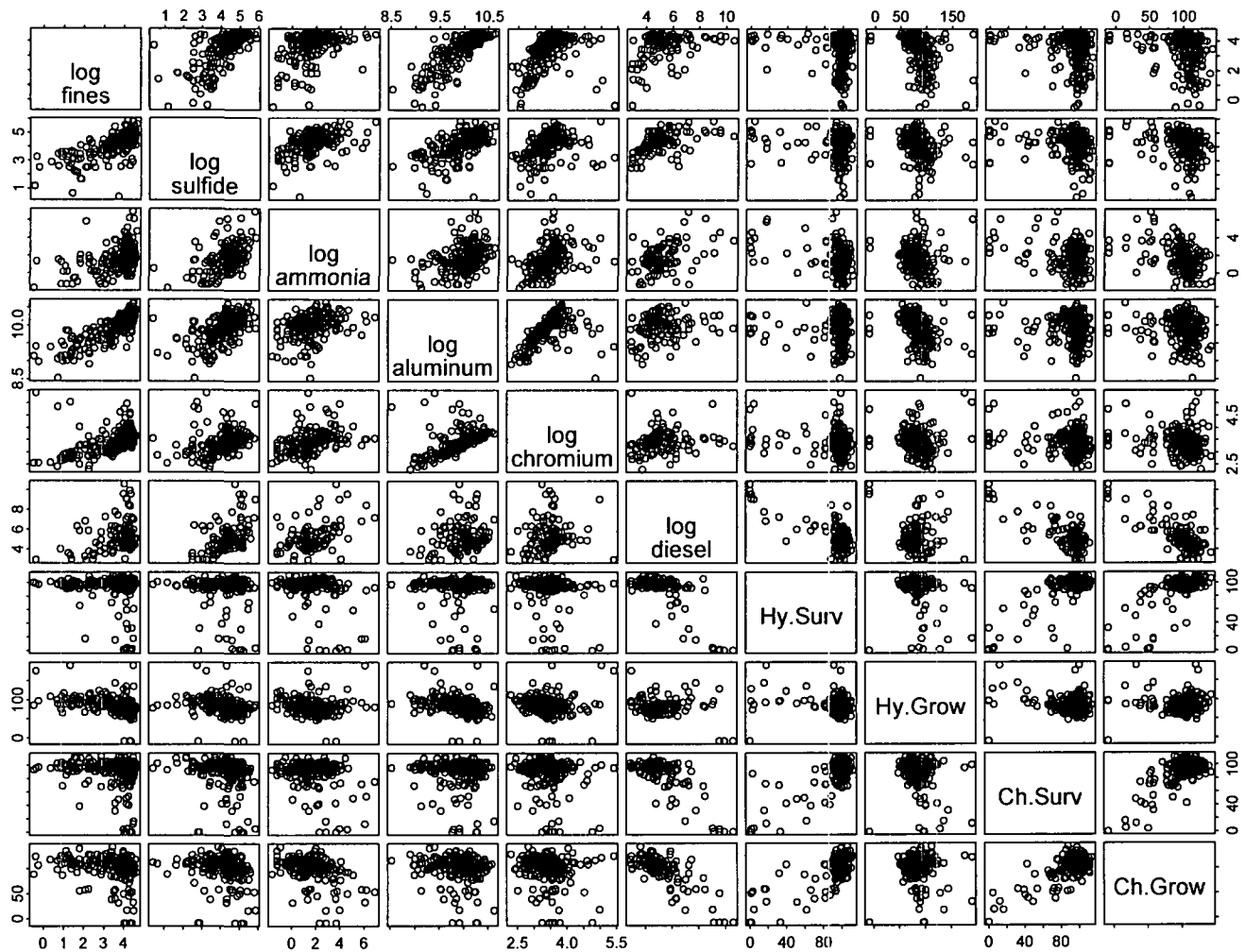


Figure 4-2. Pairwise scatter plots for metals



Biological endpoints were control adjusted; skewed chemical analytes were natural-log-transformed.

Figure 4-3. Pairwise scatter plots for biological endpoints and selected chemical and physical analytes

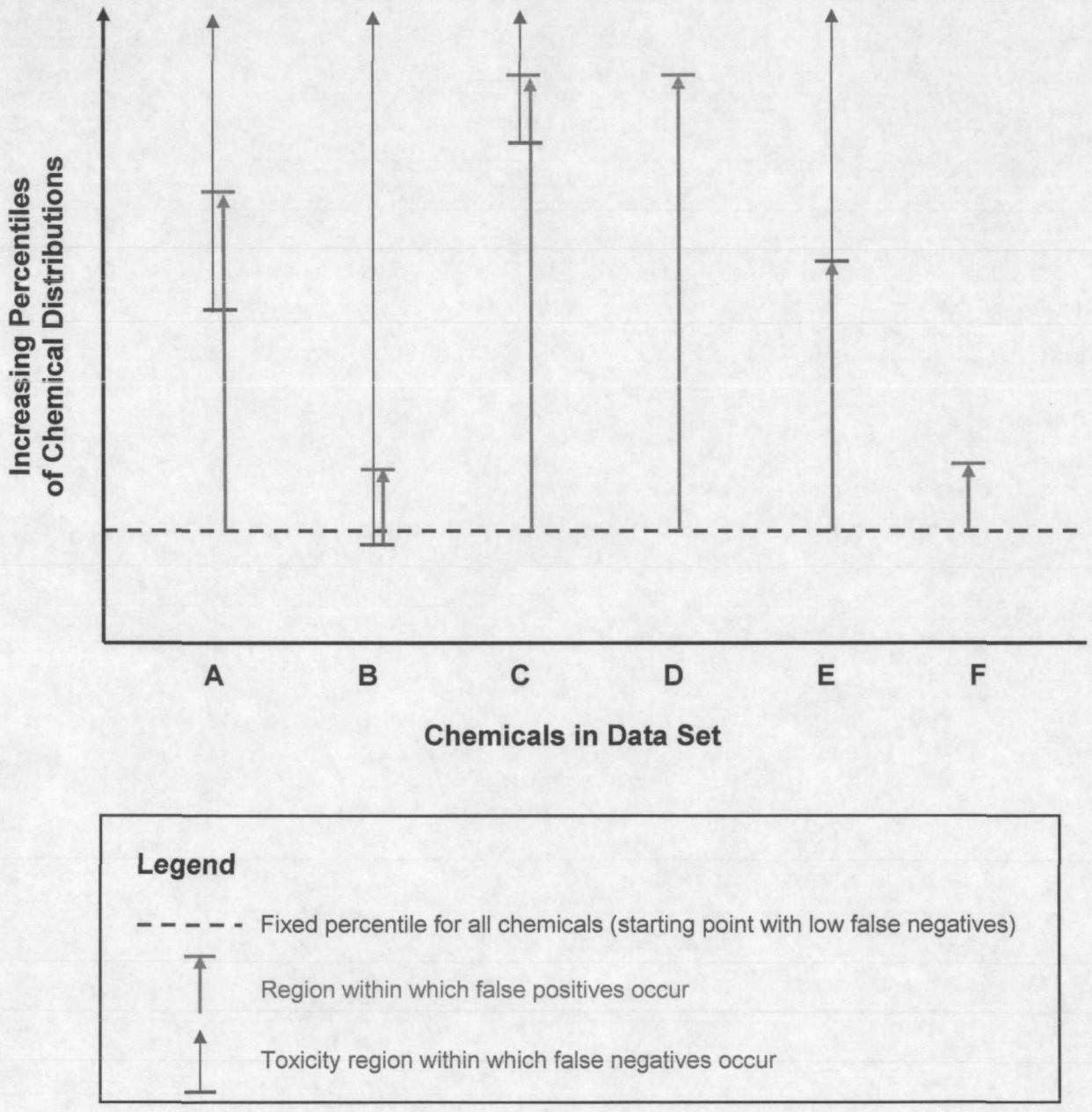


Figure 5-1. Floating percentile model

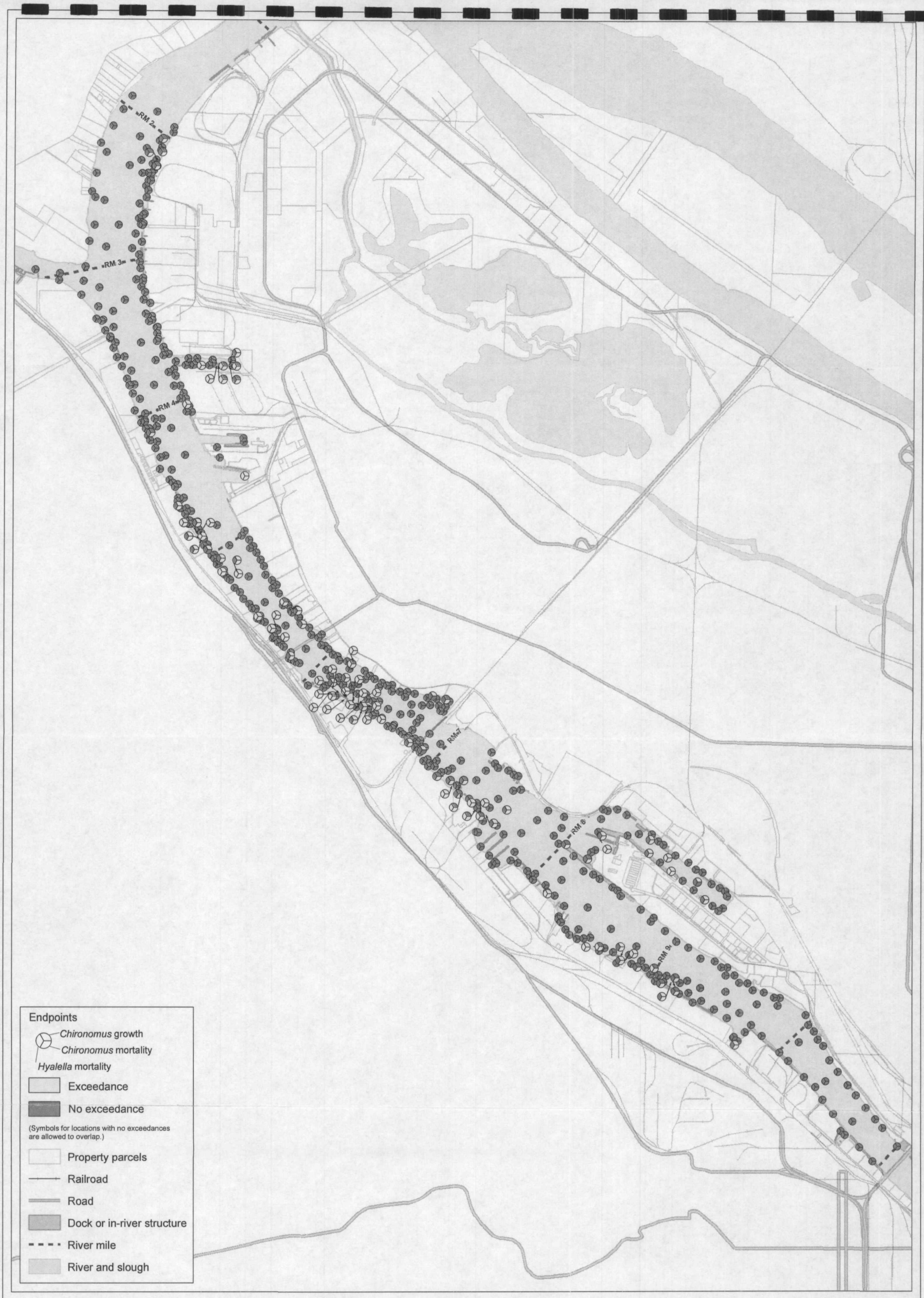


Figure 5-2. Exceedance of Effects Level 2 predicted by floating percentile model

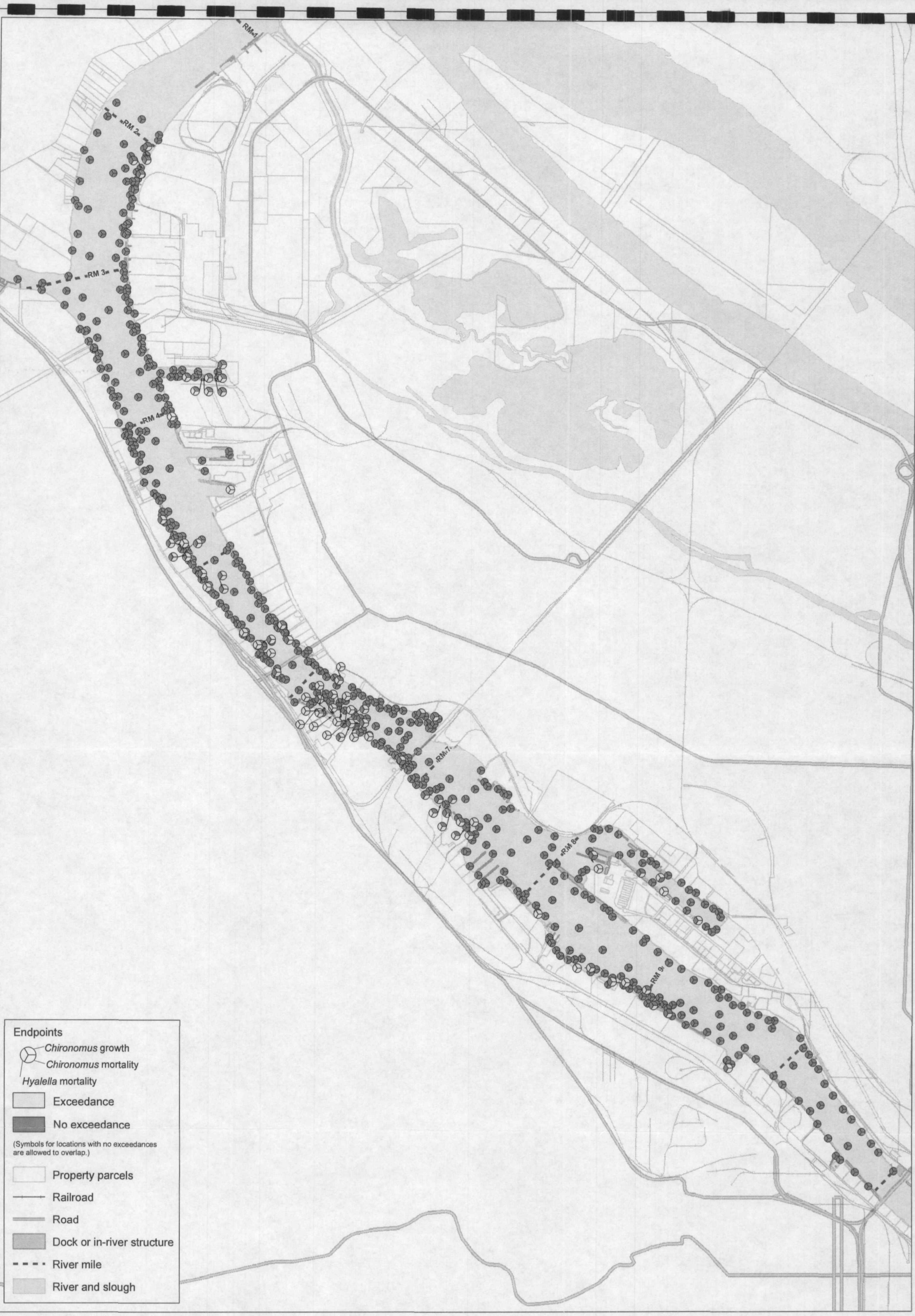


Figure 5-3. Exceedance of Effects Level 3 predicted by floating percentile model

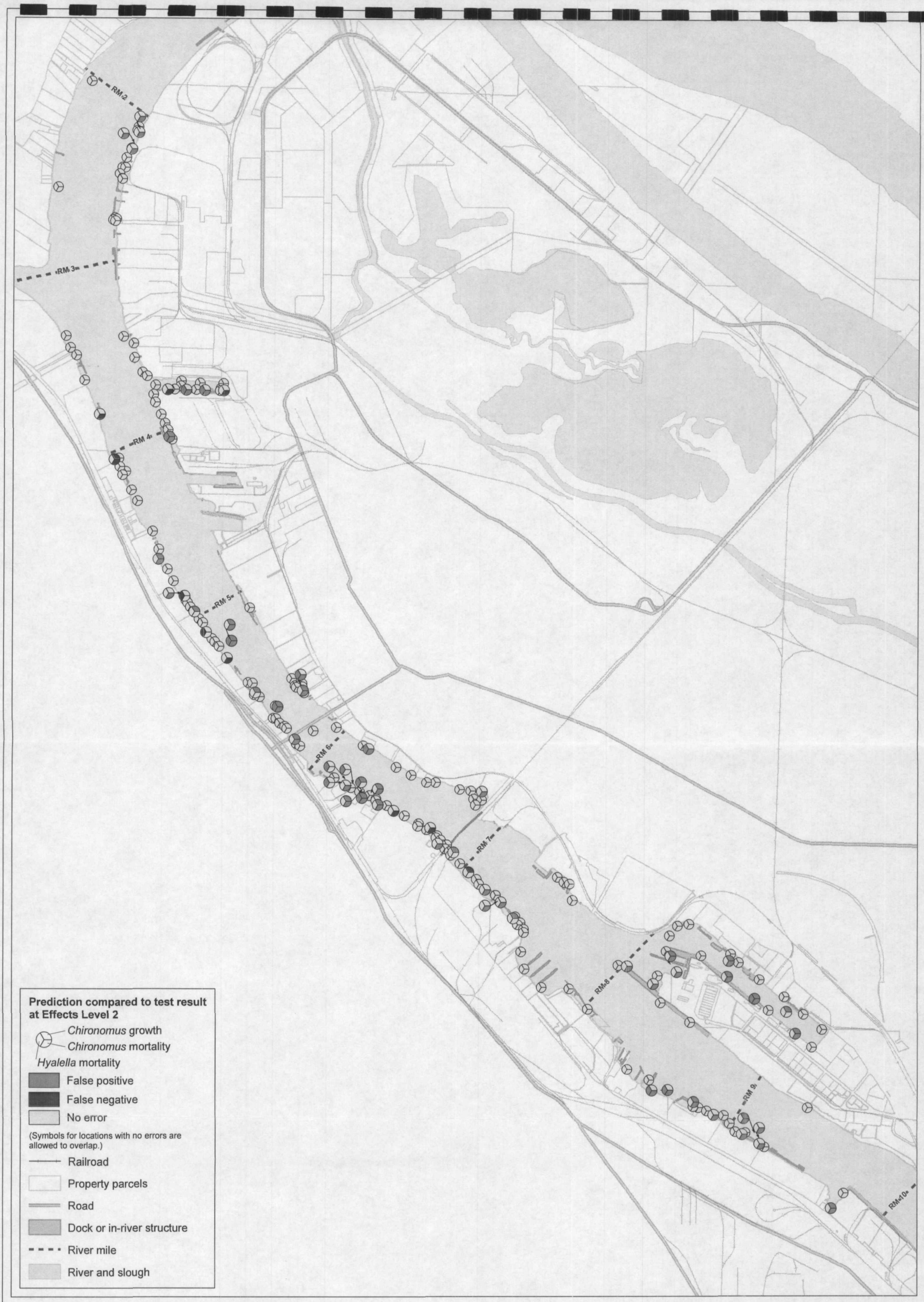


Figure 5-4. Floating percentile model false positive and false negative errors at Effects Level 2

RM 2
RM 3
RM 4
RM 5
RM 6
RM 7
RM 8
RM 9
RM 10

Prediction compared to test result at Level 3

Chironomus growth
Chironomus mortality
Hyalella mortality

False positive
False negative
No error

(Symbols for locations with no errors are allowed to overlap.)

Railroad
Property parcels
Road
Dock or in-river structure
River mile
River and slough

Figure 5-5. Floating percentile model false positive and false negative errors at Effects Level 3

0 0.25 0.5 1 Miles

0 0.5 1 2 Kilometers

N

Wind Ward environmental LLC

prepared by CJ 01-31-06 map 2171

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\FPct_Model

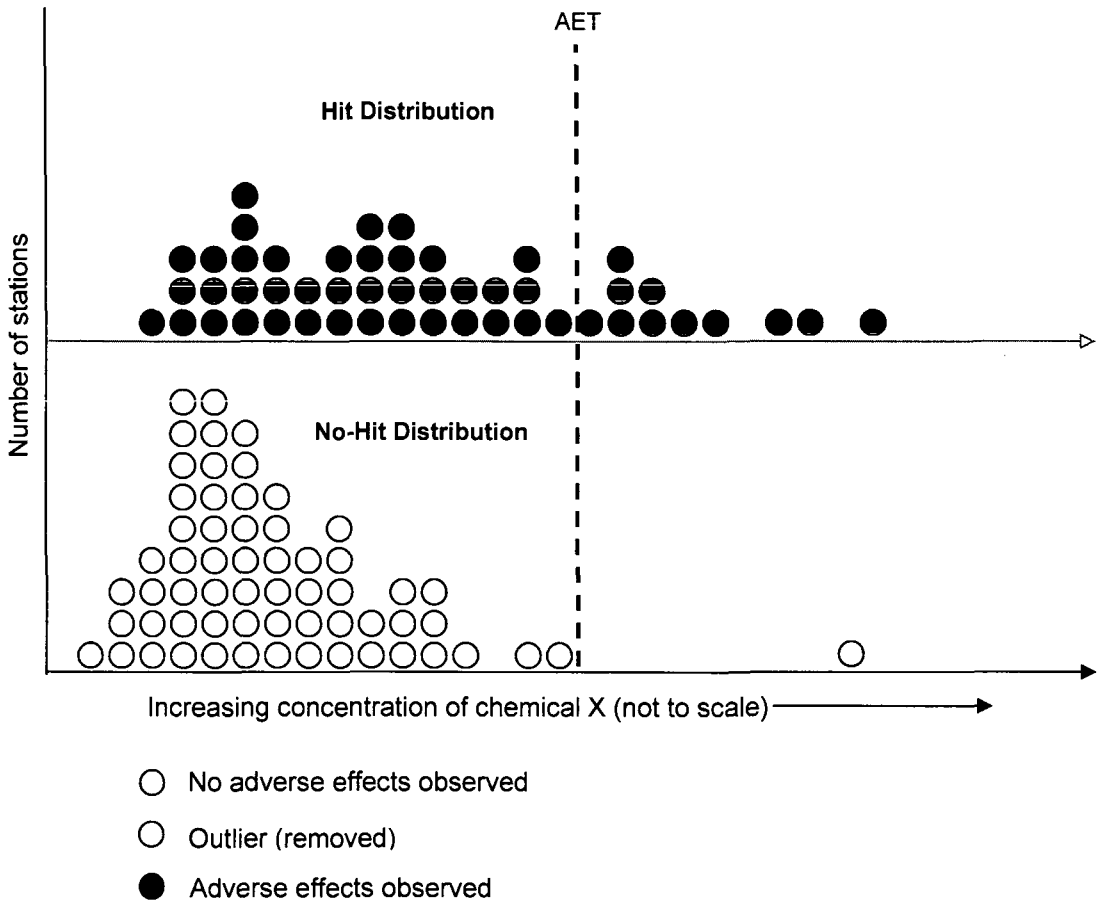


Figure 5-6. Calculation of apparent effects thresholds

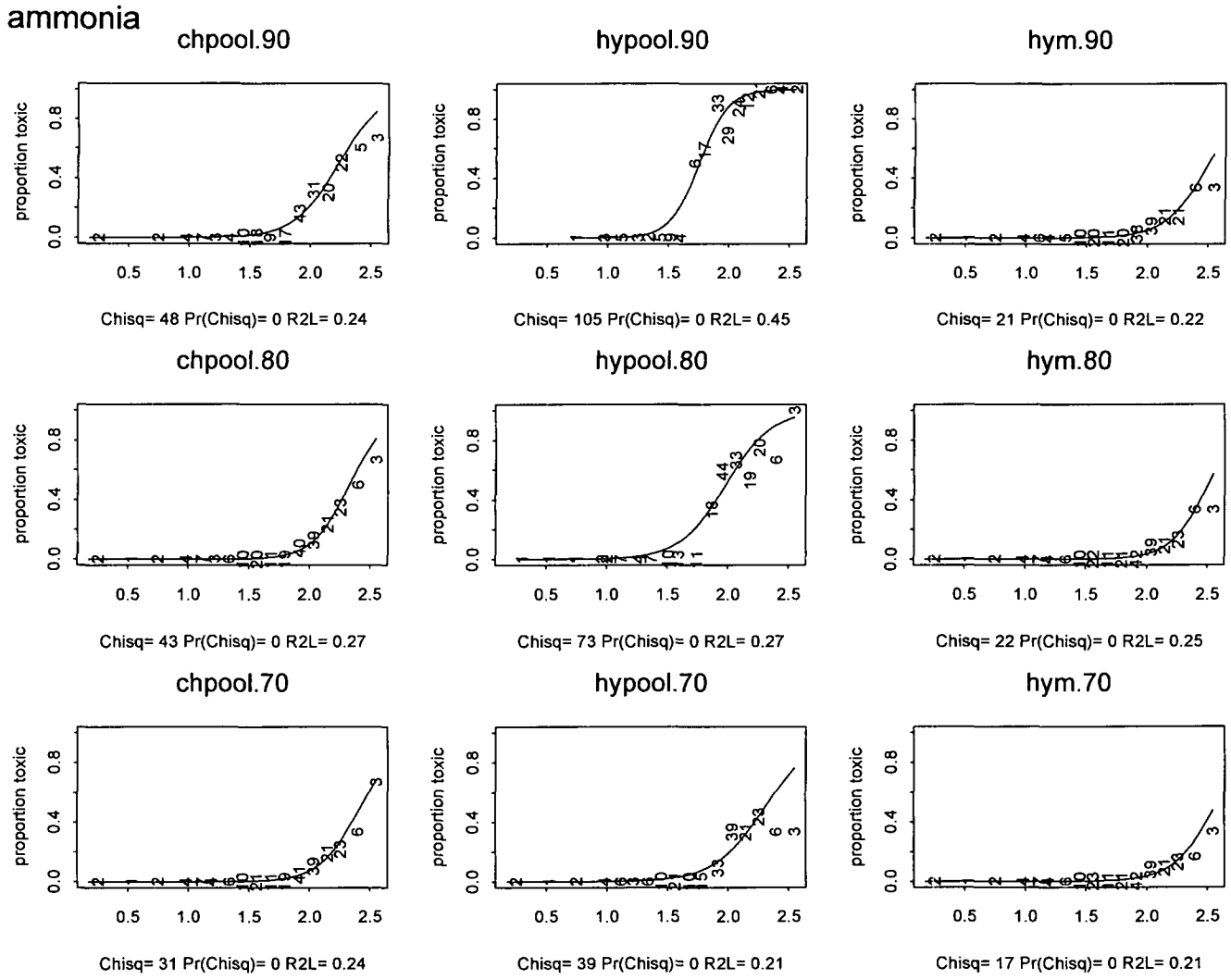
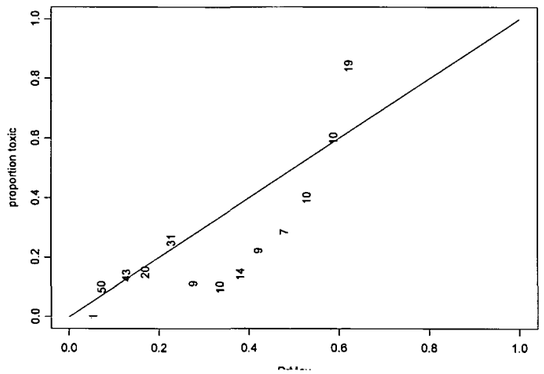
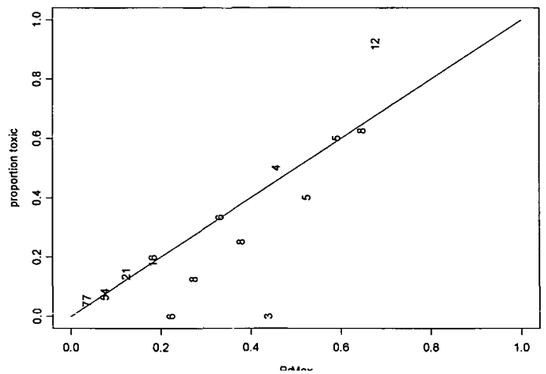


Figure 5-7. Individual logistic regression models for ammonia for three biological endpoints at three effects levels

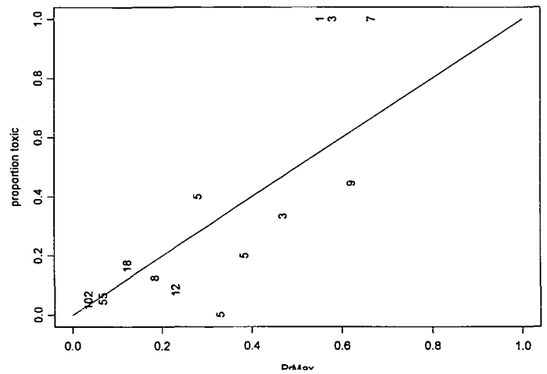
Level 1 *Chironomus* Pooled



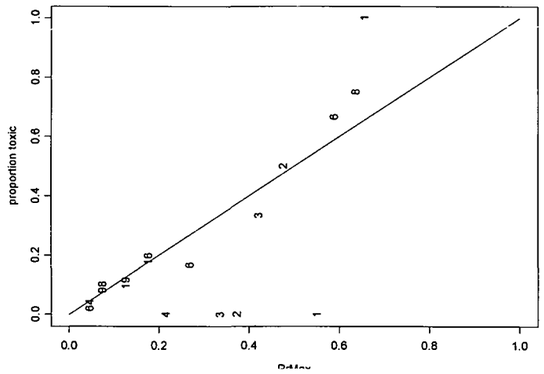
Level 2 *Chironomus* Pooled



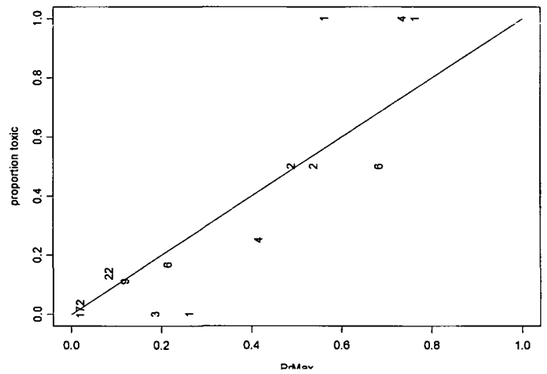
Level 3 *Chironomus* Pooled



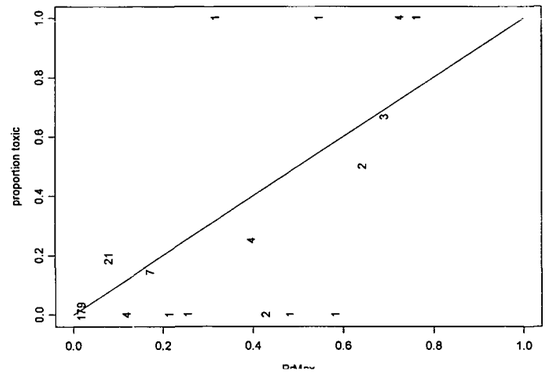
Level 1 *Hyalella* Mortality



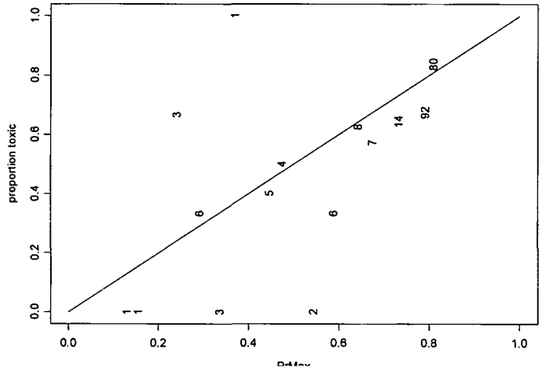
Level 2 *Hyalella* Mortality



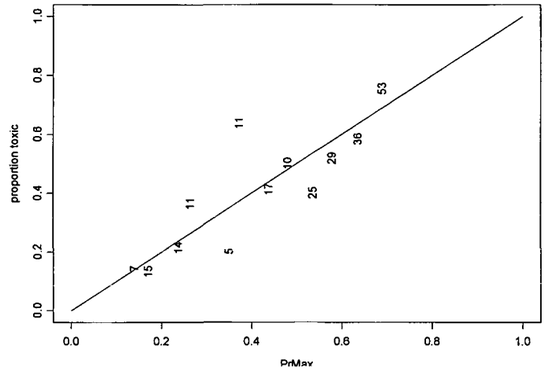
Level 3 *Hyalella* Mortality



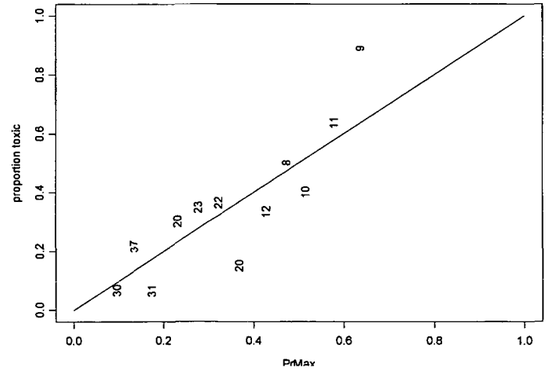
Level 1 *Hyalella* Pooled



Level 2 *Hyalella* Pooled

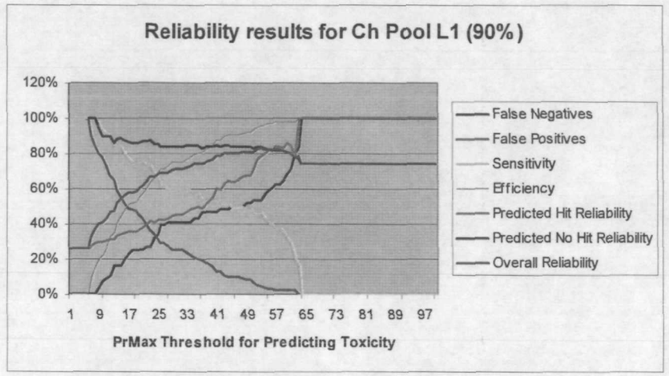


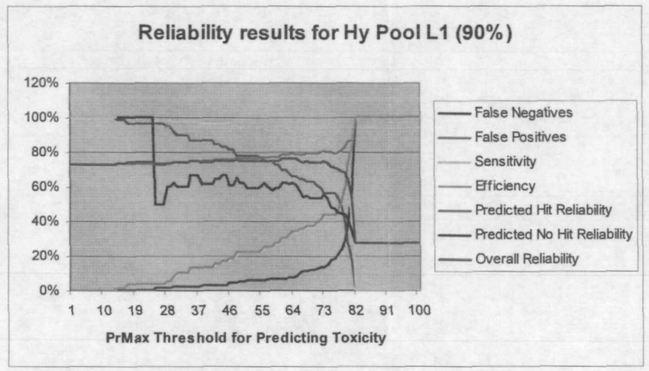
Level 3 *Hyalella* Pooled

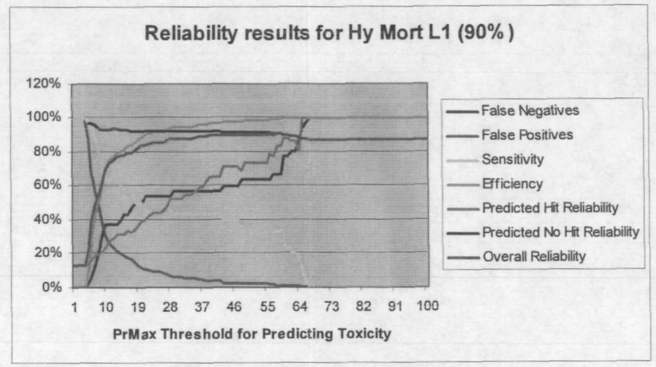


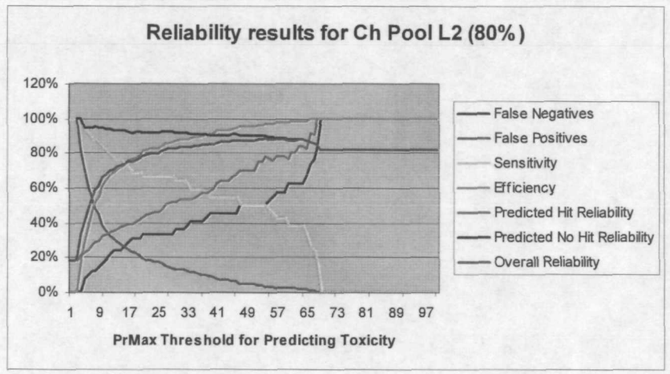
Sample sizes within each PrMax interval are shown as the plotting symbol.

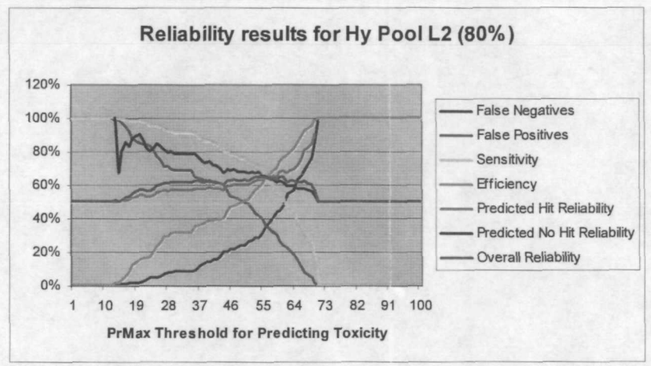
Figure 5-8. Median PrMax value and observed proportion of toxic data within PrMax intervals of 0.05

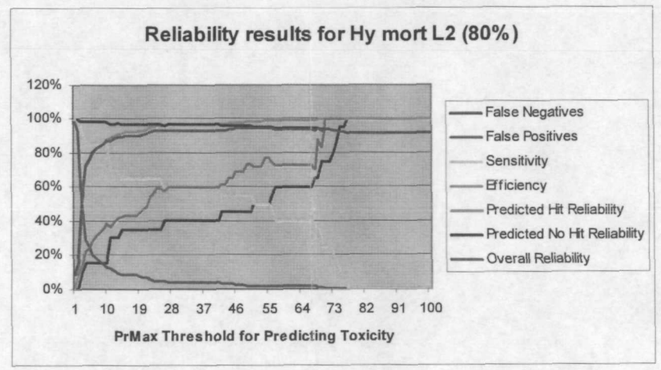


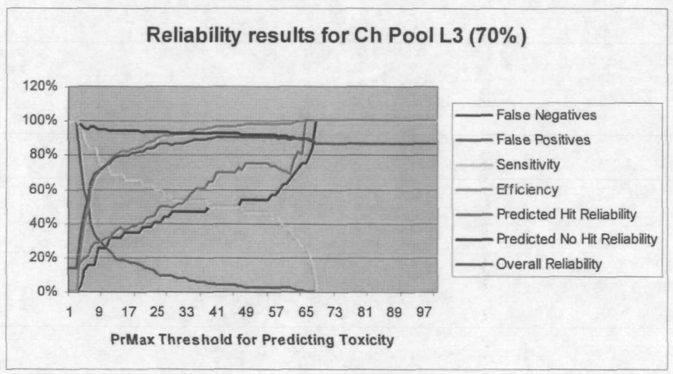


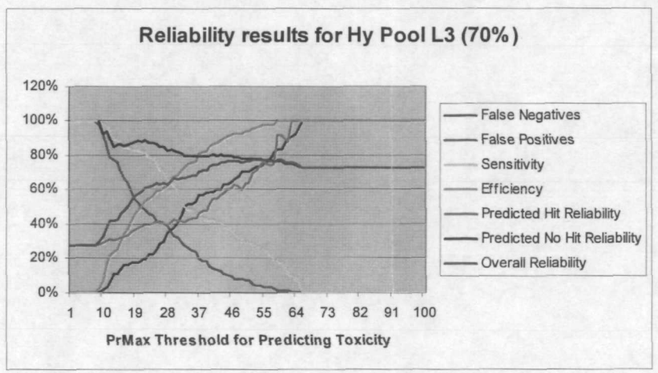












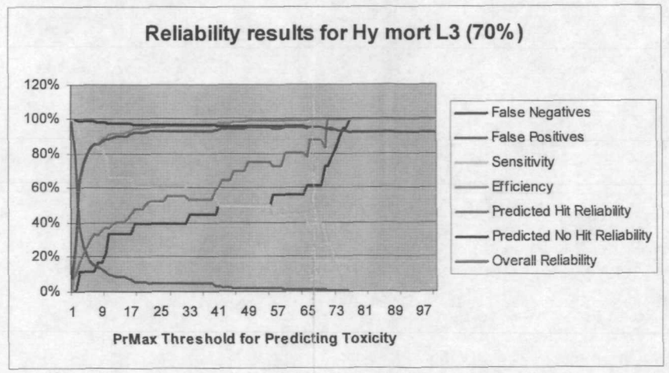


Figure 5-9. Reliability results for three biological endpoints at three effects levels

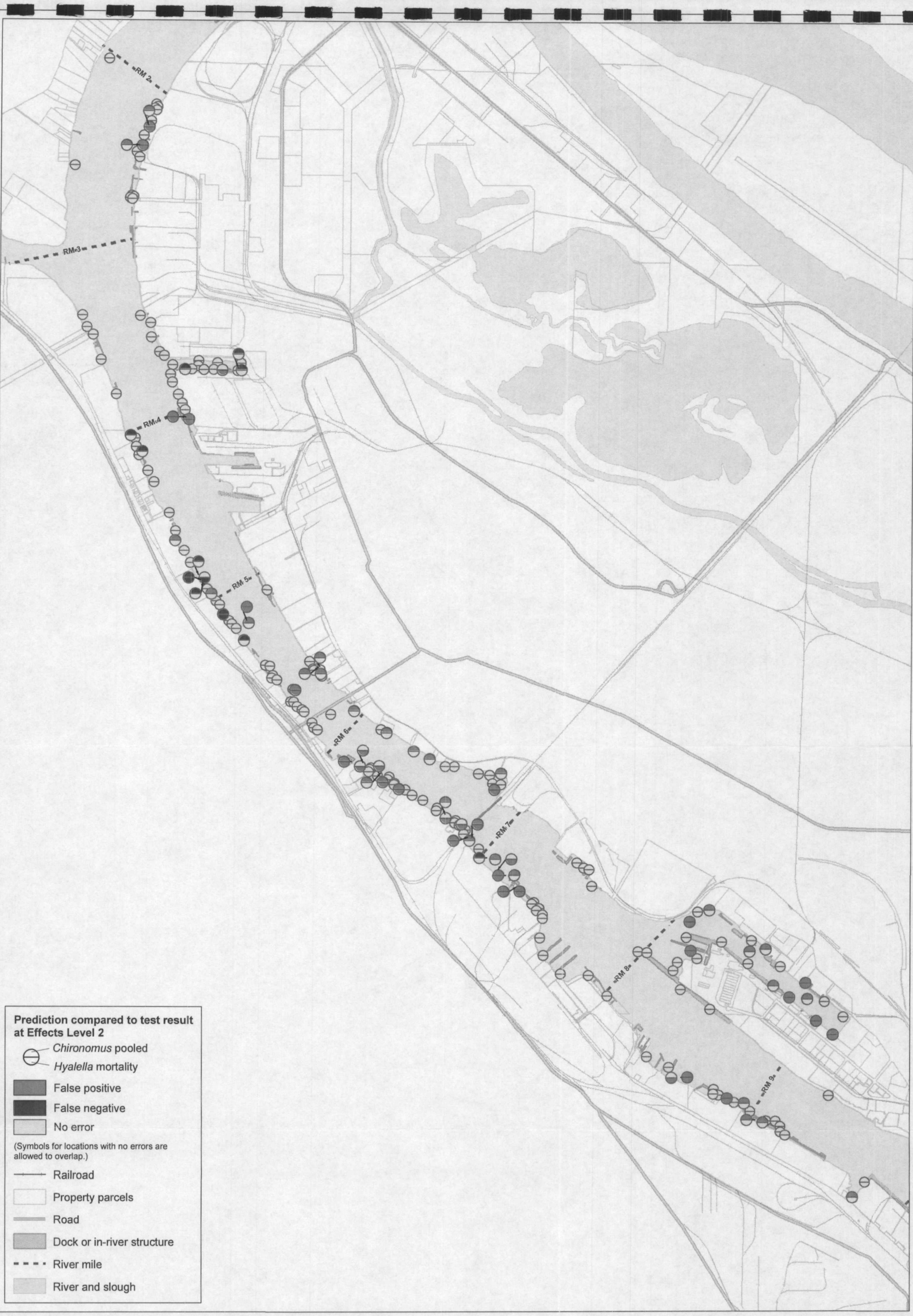


Figure 5-10. Logistic regression model false positive and false negative errors at Effects Level 2

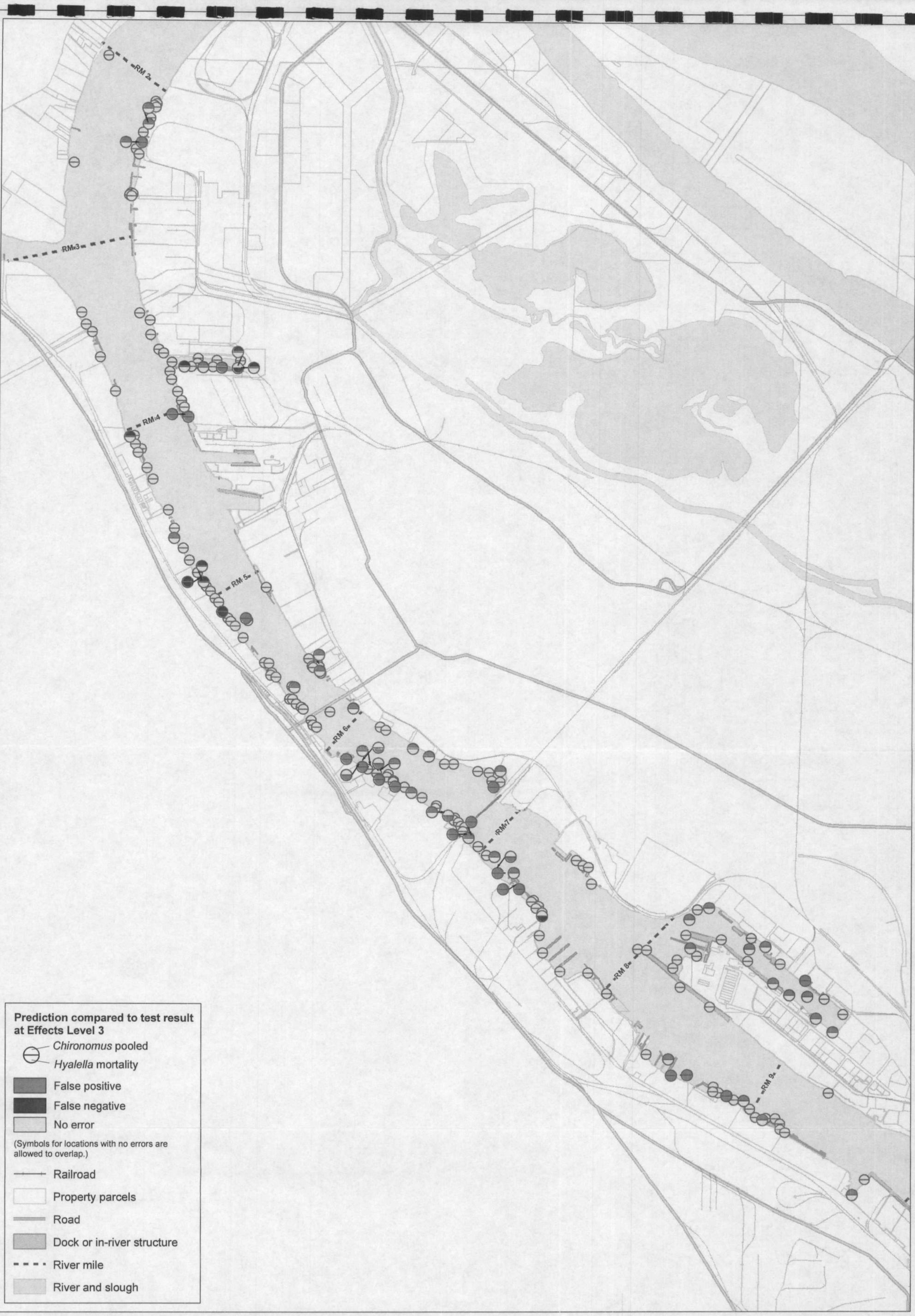
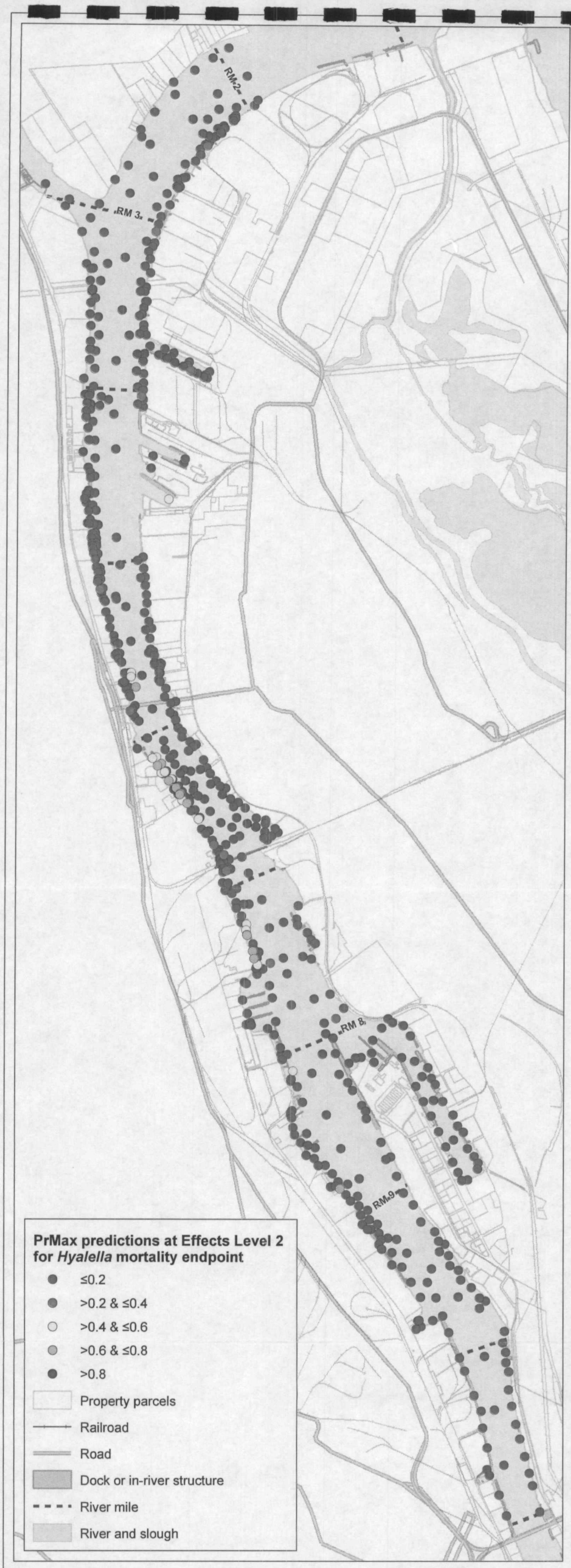


Figure 5-11. Logistic regression model false positive and false negative errors at Effects Level 3



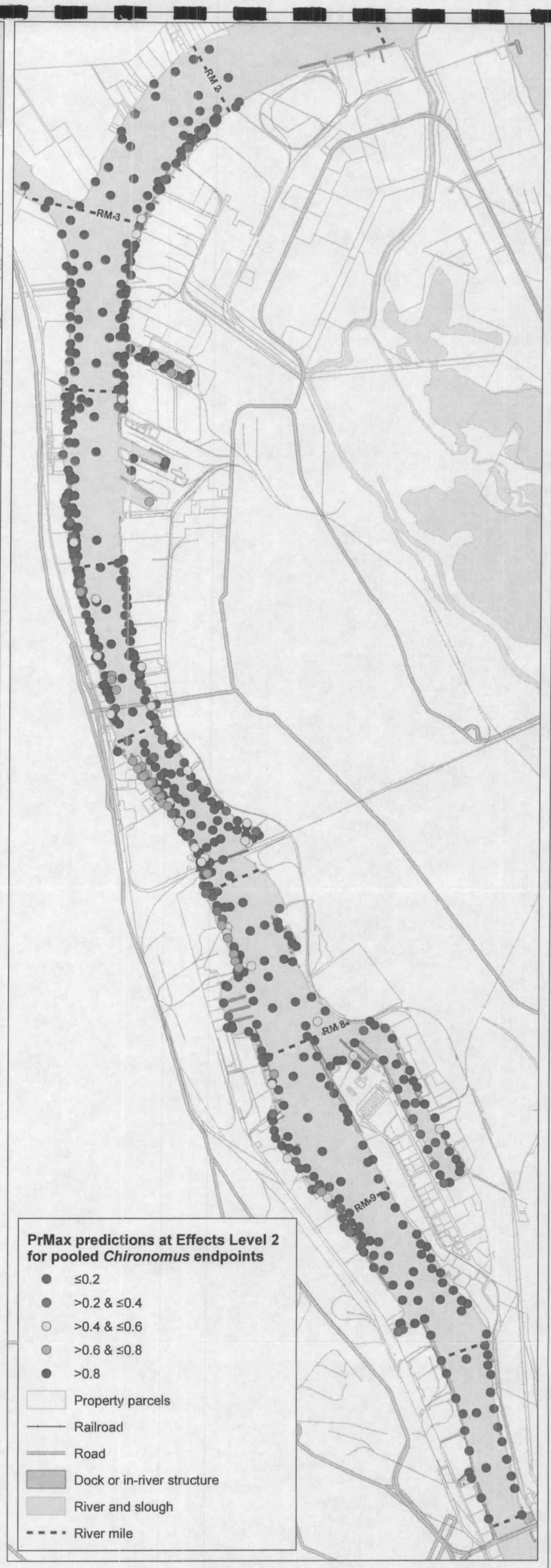


Figure 5-12. Probability of toxicity predicted for surface sediment sampling stations at Effects Level 2 using PrMax from logistic regression models

0 0.25 0.5 1 1.5 Miles

0 0.5 1 2 Kilometers

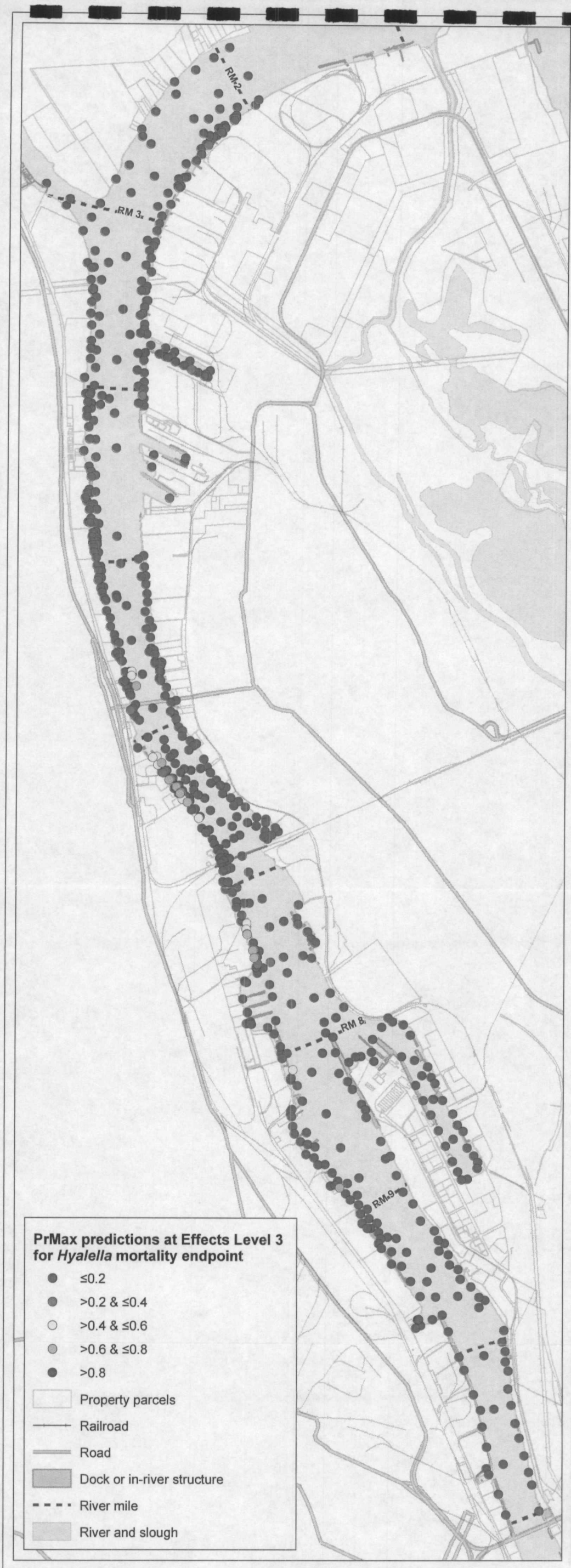
N

Wind Ward environmental LLC

Prepared by LSM 01-31-06 map 2145

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\LR_tox_models\

LRM results based on complete chemical list, including fines as individual predictor; none excluded to reduce errors. Models run and fitted 01/31/06. There are no PrMax values > 0.8.



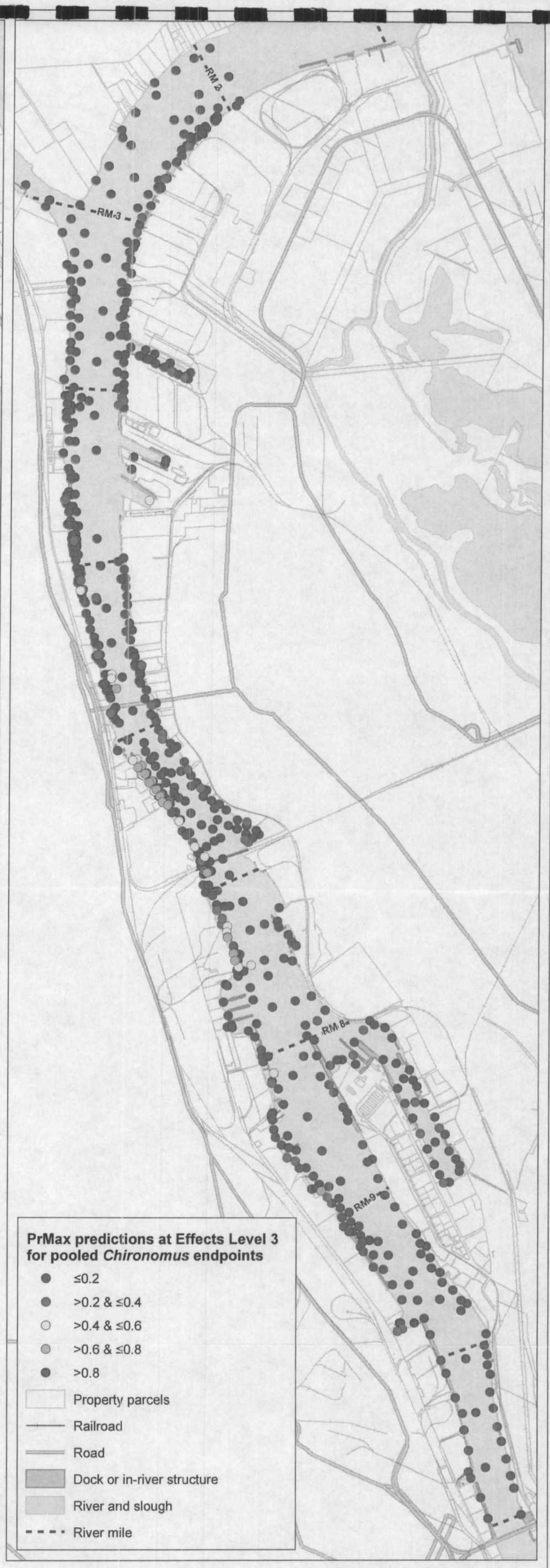


Figure 5-13. Probability of toxicity predicted for surface sediment sampling stations at Effects Level 3 using PrMax from logistic regression models

0 0.25 0.5 1 1.5 Miles

0 0.5 1 2 Kilometers

N

Wind Ward environmental LLC

Prepared by LSM 01-31-06 map 2144

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\LR_tox_models\

LRM results based on complete chemical list, including fines as individual predictor; none excluded to reduce errors. Models run and fitted 01/31/06. There are no PrMax values > 0.8.

APPENDIX A. EVALUATION OF EXISTING SQV SETS INCLUDING CHEMICAL DATA SETS

Note: Appendix A data files can be found on the accompanying compact disk.

APPENDIX A. EVALUATION OF EXISTING SQV SETS INCLUDING CHEMICAL DATA SETS

A.1 RELIABILITY OF EXISTING SEDIMENT QUALITY VALUES

This appendix provides a detailed description of the methods and results of the reliability analysis for existing SQV sets in North America.

Five SQV sets already in use in North America were included in the reliability analysis (for a more complete description of the SQV sets evaluated, see Avocet and SAIC (2002) or the specific references cited below):

- **TELs/PELs** – TELs/PELs are derived using the database percentile method. TELs are intended to represent chemical concentrations below which biological effects rarely occur. PELs are intended to represent chemical concentrations above which adverse biological effects frequently occur. TELs/PELs were derived by classifying sediment samples within each data set as either toxic or non-toxic. TELs were calculated as the geometric mean of the 15th percentile of the effects distribution and the 50th percentile of the no-effects distribution. PELs were calculated as the geometric mean of the 50th percentile of the effects distribution and the 85th percentile of the no-effects distribution. TEL/PEL values have been developed for 8 metals, 12 individual PAHs, total PCBs, and 7 chlorinated pesticides (CCME 2002).
- **TECs/PECs** – Consensus-based SQVs have been proposed by a group of private and agency sediment researchers in an attempt to unify the wide variety of SQVs available in the literature (Ingersoll et al. 2000; MacDonald et al. 2000). Threshold effects concentrations (TECs) were derived using a group of existing freshwater SQV sets that represented levels below which adverse effects were seldom observed. TECs are considered conservative screening tools and not intended for use as cleanup goals. Similarly, probable effects concentrations (PECs) were derived using a group of existing freshwater SQV sets that represented levels above which adverse effects would be expected. If three or more published values with a similar narrative intent were available for a chemical or group of chemicals, the TEC or PEC was calculated as the geometric mean of these values. TECs and PECs have been developed for 8 metals, 10 individual PAHs, total PAHs, total PCBs, and 9 chlorinated pesticides (MacDonald et al. 2000).
- **LELs/SELs** – The screening level concentration approach was developed by the Ontario Ministry of the Environment and is based on the presence or absence of benthic species in freshwater sediments (Persaud et al. 1993). First, a field database of synoptic chemical and

benthic community data was compiled. A chemical concentration distribution was prepared for each benthic species and each chemical using only those stations at which each species was observed. For each distribution, the 90th percentile was determined. This concentration is assumed to represent a conservative estimate of the upper tolerance level for that species and that chemical since above that level the species is seldom observed. For each chemical, the tolerance levels of all the species are plotted on a graph by increasing concentration. From this distribution, various levels can be selected, depending on what percent of the species is to be protected. The most widely used values, developed by the Ontario Ministry of the Environment for use in the Great Lakes, include the "lowest effect level" (5th percentile) and the "severe effect level" (95th percentile). The LEL corresponds to a level at which you would expect to see effects in only 5% of benthic species, while the SEL represents a level at which you would expect to see effects in 95% of benthic species.

- **Washington Freshwater SQS/CSL** – The floating percentile method was developed in an effort to improve the reliability of freshwater SQVs for Washington State (Avocet 2003; Avocet and SAIC 2002). An optimal percentile of the data set that provides a low false negative rate is selected, and then each individual chemical concentration is adjusted upward until the false positive rate has decreased to its lowest possible level while retaining the same false negative rate. The method is designed to reduce mathematical error associated with the use of fixed percentiles for all chemicals. Sediment quality standards (SQS) and cleanup screening levels (CSLs) were calculated using the FPM for 11 metals, 16 individual PAHs, LPAHs, HPAHs, 4 phthalates, dibenzofuran, and total PCBs. These SQVs were derived using a large data set, primarily from western Washington and Oregon and including all of the Portland Harbor data that existed at that time (2001), and are currently applicable to freshwater sediments in Washington State (Avocet 2003).
- **Quotient Methods** – Quotient methods were developed as an approach to increase the predictive ability of certain SQVs described above (Long et al. 1998), and have been applied to TELs/PELs and TECs/PECs. Several quotient methods are available, some of which use individual metals and PAHs and others of which sum chemical classes. Based on the exploratory analysis conducted for this data set, several chemical classes such as PAHs and PCBs appeared to be more predictive of toxicity when summed. Therefore, quotients that use summed values, such as the mean PEL-Q, may be more appropriate. This is also the approach recently adopted for use in British Columbia (Macfarlane et al. 2002). However, it does not include all of the chemicals of interest at the site. Therefore, an

alternative version was also evaluated (SQG-Q) based on a recent paper by Fairey et al. (2001), which includes additional chemicals of interest, such as chlordanes and dieldrin.

For each existing SQV set, the more protective of the two thresholds (TEL, TEC, LEL, and SQS) was compared to the Level 1 and 2 biological effects levels, and the higher of the two thresholds (PEL, PEC, SEL, and CSL) was compared to the Level 3 biological effects levels, consistent with the narrative intent of these SQVs.

A.2 RELIABILITY ASSESSMENT METHODS

This section presents the methods used to obtain the appropriate chemistry data for the comparison of each SQV set and to evaluate the toxicity test endpoints. The chemistry data methods are presented in Section B.2.1; the toxicity data methods are presented in Section B.2.2.

A.2.1 Chemistry data methods

The project database was queried to obtain all chemistry data for the selected group of analytes (depending on the SQV set being evaluated), excluding any data qualified with a U, N, or R (see Section 2.2.1). To evaluate the reliability of existing SQV sets, chemical concentrations were summed in the same manner as that used in deriving each set of existing SQVs (e.g., threshold effects levels [TELs] and probable effects levels [PELs]) to facilitate comparison. For example, if the SQV set included values for individual PAHs, individual PAH concentrations were used in the reliability analysis. If the SQV set used low-molecular-weight PAH (LPAH) and high-molecular-weight (HPAH) sums, these sums were used instead.

These data were downloaded into Microsoft Excel® files, which are included in this appendix. There are 15 Excel® files, one for each combination of the three effects levels and five endpoints (four individual endpoints and one pooled endpoint). For SQV sets other than the PEL-Qs, the following approach was used. The first worksheet, entitled "BioHits," contains the biological hit/no-hit results for the endpoint and effects level being evaluated. The worksheet "ChemData" shows the chemistry data for all stations downloaded from the SEDQUAL Information System, organized by chemical and increasing concentration. A Visual Basic® macro called MakeTable is then run to organize the data into a data table, shown in the worksheet DataTable. The DataTable worksheet also has a column into which the biological hit/no-hit values are entered for each station. Blank cells indicate analytes for which no data are available at those stations. The reliability macro skips these cells.

The final worksheet, entitled Criteria, contains the individual SQVs for each of the four SQV sets that are being assessed for the 34 analytes included among the various SQV

sets.[1] These values are pre-entered in columns H-AO of the worksheet. To the left of these values, there are columns for each of the seven measures of reliability, which are calculated by a Visual Basic® macro called TestReliability. The TestReliability macro compares the chemical concentrations of each chemical at a station to the corresponding SQVs and determines whether a hit or no-hit would be predicted at that station. Then the chemical hit/no-hit prediction is compared to the biological hit/no-hit value, and the macro records whether the result is a correct prediction, a false positive, or a false negative. From these results, each of the other reliability parameters was calculated. These and the other Excel® macros were manually verified to ensure their accuracy. The seven reliability parameters are listed below:

- **False negatives** – Incorrectly predicted no-hits/total hits
- **False positives** – Incorrectly predicted hits/total no-hits
- **Sensitivity** – Correctly predicted hits/total hits
- **Efficiency** – Correctly predicted no-hits/total no-hits
- **Predicted hit reliability** – Correctly predicted hits/total predicted hits (this measure is equivalent to "1988 Efficiency" in Avocet (Avocet 2003; Avocet and SAIC 2002))
- **Predicted no-hit reliability** – Correctly predicted no-hits/total predicted no-hits
- **Overall reliability** – Correctly predicted stations/total stations

For the quotient methods, the chemistry was downloaded, and both the probable effects level quotient (PEL-Q) and the sediment quality guideline quotient (SQG-Q) were calculated for each station. The PEL-Q was calculated for each sediment sample by summing the average quotient for seven metals (i.e., arsenic, cadmium, chromium, copper, lead, mercury, and zinc), the quotient for total PAHs, and the quotient for total PCBs and then dividing this sum by three. The SQG-Q used the sum of the quotients of each individual chemical or class included in the equation, divided by the number of chemicals or classes, and was calibrated using an empirical approach in which a variety of different equations was tested using various possible SQGs as the basis for the quotient. The chemicals included, and the SQGs on which their quotients are based, are: cadmium (PEL), copper (effects range median [ERM]), silver (PEL), lead (PEL), zinc (ERM), total chlordane (ERM), dieldrin (ERM), total PAHs (PEC), and total PCBs (PEC). PAHs are also OC-normalized in this approach.

A.2.2 Toxicity data methods

Two endpoints, growth and mortality, were included in the reliability assessment. The mortality endpoint was obtained for both toxicity tests at all 233 stations, whereas the growth endpoint could not be obtained for a few stations because of 100% mortality in

[1] The macros for the spreadsheets were set up using the word "criteria." However, for the Portland Harbor project, the word "criteria" should be replaced with the word "SQV."

the same samples. The types and numbers of toxicity test endpoints in the Round 2 data set are summarized in Table A-1.

Table A-1. Round 2 toxicity tests and endpoints

Test	Maximum Number of Stations[a]
Hyalella azteca	
28-day mortality	233
28-day growth	229
Chironomus tentans	
10-day mortality	233
10-day growth	227

[a] Some of the stations may have been labeled "Indeterminate" for one or more of the effects levels. The number of endpoints directly correlates to the number of stations.

For the reliability assessment, each of the four individual endpoints was assigned to the three biological effects levels based on the definitions stated in Section 2.2.3. In addition, a pooled endpoint was derived by combining all four endpoints from the two tests. Table A-2 shows the number and percentage of stations associated with biological hits for each effects level and endpoint combination.

Table A-2. Biological hits

Effects Level	Number of Biological Hits (percent)[a]				
	Chironomus growth	*Chironomus* mortality	*Hyalella* growth	*Hyalella* mortality	Pooled endpoint[b]
Level 1	29 (13%) [12]	47 (21%) [11]	139 (66%) [18]	30 (13%) [3]	167 (78%) [18]
Level 2	24 (11%) [0]	34 (15%) [0]	98 (43%) [0]	20 (9%) [0]	128 (55%) [0]
Level 3	17 (7%) [0]	25 (11%) [0]	46 (20%) [0]	18 (8%) [0]	77 (33%) [0]

[a] The denominator used to determine the percentage of hits excludes the number of statistically indeterminate samples shown in brackets.

[b] For this analysis, all four biological endpoints were combined into a single pooled endpoint. For later analyses, biological endpoints were pooled by species.

As can be noted from Table A-2, there were substantial differences among endpoints in the observed responses. The *Hyalella* growth test showed a response at a greater number of stations than any of the other toxicity test endpoints for all effects levels. The *Chironomus* growth test was comparable to the *Hyalella* mortality test in the number of adverse responses exhibited at each effects level; they both exhibited the fewest number of responses among the endpoints. *Chironomus* mortality was intermediate in the number of responses exhibited at each effects level. The pooled endpoint always

exhibited a response at a relatively large number of stations as compared to any one individual endpoint, suggesting that there were frequent differences in the endpoints exhibiting effects among stations.

A.3 RELIABILITY ANALYSIS

The reliability analysis for each of the effects levels is discussed in this section. To simplify the discussion, the evaluation below focuses on the four primary reliability parameters: sensitivity, efficiency, predicted no-hit reliability, and predicted hit reliability. Two of the other parameters, false positives and false negatives, are simply 100% minus sensitivity and efficiency. The final parameter, overall reliability, is less useful in this analysis because it is dependent on the proportion of hits to no-hits in the data set, which varies significantly among effects levels.

A.3.1 Level 1

Table A-3 presents the results for the four SQV sets that were assessed at Level 1. The TEL, TEC, and LEL levels all performed similarly and very conservatively, although in general, the TECs performed 10 to 15% better with respect to efficiency than the TELs and LELs. In all three cases, the SQV sets had very high sensitivity (few false negatives). On the other hand, these SQV sets classified nearly every sample as a hit, leading to a very high false positive rate (100% in the case of the TELs). In general, these SQV sets predicted that all or nearly all samples would be hits, and the proportion of correctly predicted hits simply reflects the proportion of actual biological hits in the data set. Therefore, these SQV sets are not really useful in making correct predictions about lower effects levels. Although it is highly likely that any sample with chemical concentrations that fall below these levels will not exhibit biological effects, there will be few to no samples with chemical concentrations that are that low. Relatively large, apparent variations in the predicted no-hit reliability parameter actually represent only a few samples, inasmuch as very few samples overall are predicted to be no-hits.

Table A-3. Reliability analysis for Level 1 biological effects

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
***Chironomus* Growth**				
TEL	100	10	13	100
TEC	100	23	14	100
LEL	97	10	12	67
Washington SQS	83	51	17	91
***Chironomus* Mortality**				
TEL	98	7	20	67
TEC	94	20	22	90
LEL	96	6	20	33
Washington SQS	68	47	23	81
***Hyalella* Growth**				
TEL	98	23	59	na
TEC	88	34	60	34
LEL	99	26	60	67
Washington SQS	60	54	60	31
***Hyalella* Mortality**				
TEL	98	2	13	67
TEC	85	15	14	93
LEL	98	2	13	33
Washington SQS	57	43	15	89
Pooled Endpoint				
TEL	98	27	71	na
TEC	90	42	73	34
LEL	99	29	72	33
Washington SQS	63	61	75	23

na – did not predict any no-hits at this effects level

The Washington State freshwater SQS values are less conservative than the other three SQV sets. While they have 20 to 40% higher efficiency, it comes at the expense of 20 to 40% lower sensitivity, particularly for the more sensitive 28-day *Hyalella* endpoints, which were not included in the original calculation of these SQVs due to the lack of sufficient data at that time. These SQVs likely need to be recalculated to take into account the chronic bioassay data in order to obtain better performance with this data set.

A.3.2 Level 2

Table A-4 shows the reliability results for Level 2, which are overall very similar to those of Level 1. Again, the TEL, TEC, and LEL SQVs all classify nearly all samples as hits, resulting in high sensitivity and very low efficiency. The predicted hit and predicted no-hit reliability values appear different from those of Level 1; but in reality, these values just reflect the fact that there are fewer actual hits at Level 2, especially for

the *Hyalella* toxicity test endpoints. Therefore, the predicted hit reliability declines because most samples are still predicted to be hits. For the Washington freshwater SQS values, the same pattern is observed – sensitivity and efficiency are nearly the same as those at Level 1, while predicted hit reliability declines because there are fewer biological hits at this level, especially in the *Hyalella* test.

Table A-4. Reliability analysis for Level 2 biological effects

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
***Chironomus* Growth**				
TEL	100	4	10	100
TEC	100	17	12	100
LEL	96	4	10	67
Washington SQS	83	46	14	96
***Chironomus* Mortality**				
TEL	100	2	15	100
TEC	97	14	16	97
LEL	97	1	14	67
Washington SQS	76	43	19	91
***Hyalella* Growth**				
TEL	99	4	42	67
TEC	92	19	44	72
LEL	100	5	42	100
Washington SQS	62	45	43	61
***Hyalella* Mortality**				
TEL	100	1	9	100
TEC	100	14	10	100
LEL	95	1	8	67
Washington SQS	80	42	12	96
Pooled Endpoint				
TEL	99	2	55	67
TEC	94	20	59	72
LEL	99	2	55	67
Washington SQS	66	49	61	54

A.3.3 Level 3

The reliability results for Level 3 are presented in Table A-5. Most of the SQV sets appear to perform better at this effects level, with a few exceptions (notably a lack of sensitivity in comparison to the *Hyalella* growth results). At this level, the Washington CSLs come more into line with the other SQV sets, tending to be most similar to the PELs in performance. Among all the SQV sets, there is a better balance between sensitivity and efficiency, although judging by the low predicted hit reliability values, there is still a tendency to over-predict actual hits by a substantial amount (three times the actual number of hits).

Table A-5. Reliability analysis for Level 3 biological effects

SQV Set	% Sensitivity	% Efficiency	% Predicted Hit	% Predicted No-Hit
***Chironomus* Growth**				
PEL	82	59	13	97
PEC	65	70	14	95
SEL	53	80	16	95
Washington CSL	65	54	9	95
***Chironomus* Mortality**				
PEL	68	57	16	94
PEC	56	68	17	93
SEL	52	79	23	93
Washington CSL	72	53	16	94
***Hyalella* Growth**				
PEL	44	56	19	80
PEC	31	66	17	79
SEL	31	80	25	82
Washington CSL	51	52	20	81
***Hyalella* Mortality**				
PEL	72	56	12	96
PEC	67	68	15	96
SEL	67	79	21	97
Washington CSL	83	53	13	97
Pooled Endpoint				
PEL	57	59	40	74
PEC	45	70	42	72
SEL	41	84	55	74
Washington CSL	61	55	40	74

A.3.4 Quotient method

Pooled results for the SQG-Q and PEL-Q methods are shown in Figures A-1 and A-2, respectively. The x-axes present the full range of quotient values (SQG-Q and PEL-Q), and the y-axes present the percentage of hit classification. At each level of effects, a full range of possible quotients was evaluated to determine if there was a quotient level that could reliably predict hits and no-hits in the data set. The pink line shows the percentage of no-hits below the quotient value, while the blue line shows the percentage of hits above the quotient value. Ideally, both levels would be high (e.g., above 80%) in order for a selected quotient value to have good reliability in predicting both hits and no-hits. As can be seen from the graphs, this does not occur at any effects levels throughout the range of possible quotient values, except in some cases at the extreme ends of the data distribution. Setting values at the ends of the distributions would not be helpful because only a few stations fall below these levels (at the low end) or above these levels (at the high end).

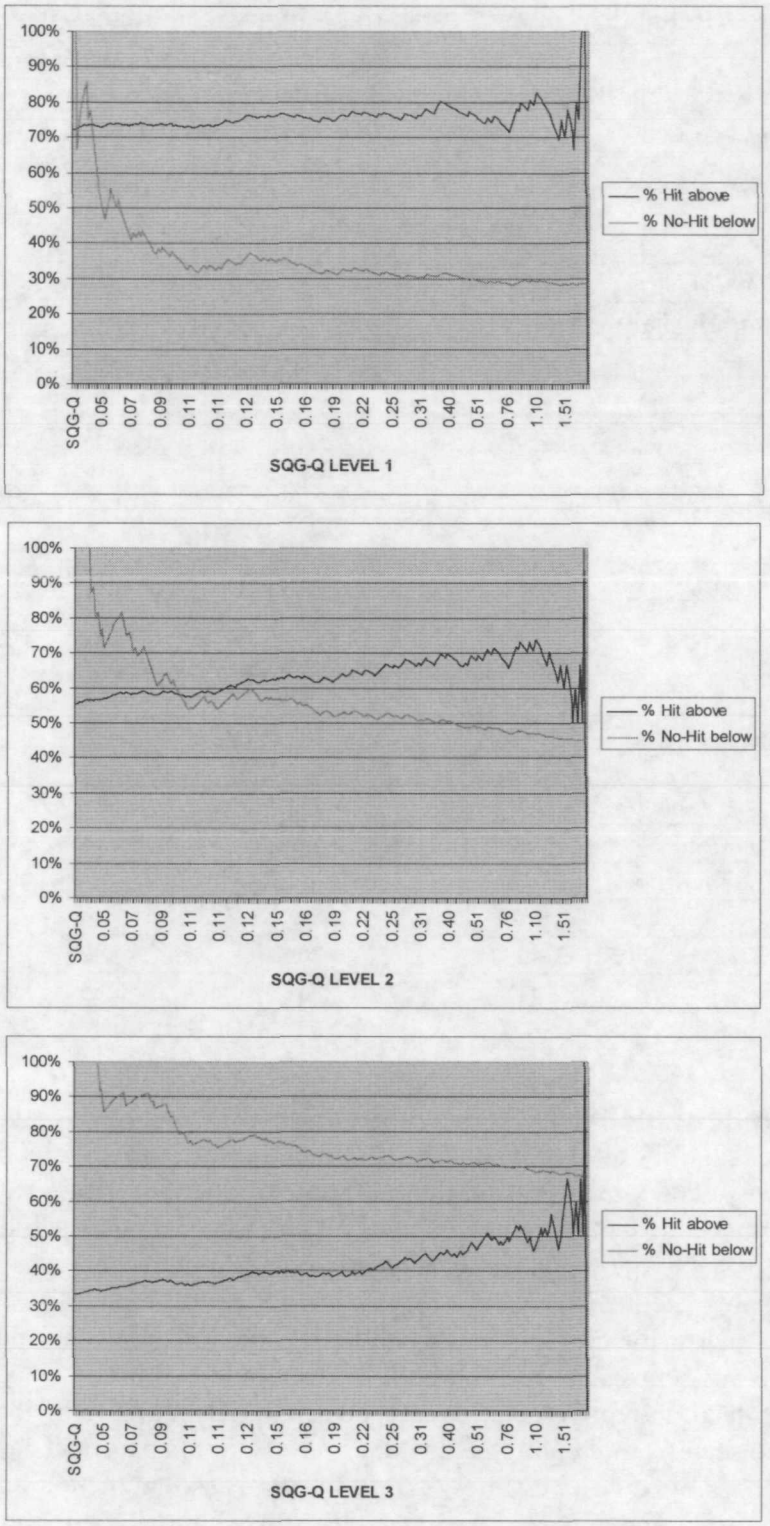


Figure A-1. SQG-Q pooled endpoint hit and no-hit screening curves

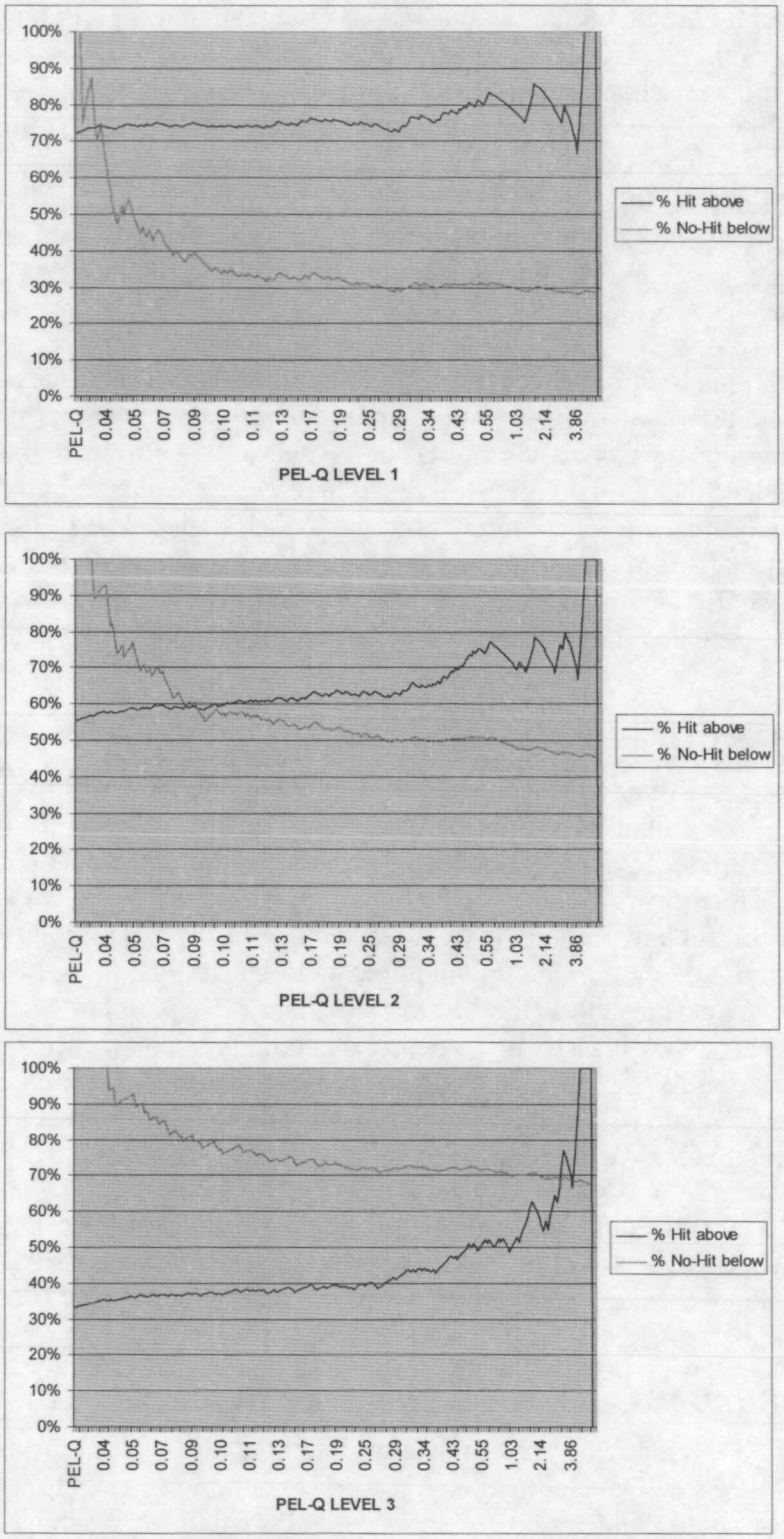


Figure A-2. PEL-Q pooled endpoint hit and no-hit screening curves

Even though a single quotient value may not be reliable for predicting both hits and no-hits, lower levels could be used to screen out areas (identify no-hits), and higher levels could be used to screen in areas (identify hits). Unfortunately, this approach also has very low reliability. At Level 1, the no-hit screening (the pink line) has a reliability of only about 30 to 40% across most of the distribution. At Level 3, the hit screening (the blue line) has only about 40% reliability through most of the data set, rising to 60% near the upper end. The intermediate Level 2 effects level has the best balance of reliability for both quotient measures but only achieves about 60% reliability for both hit and no-hit screening.

In general, this is an improvement over most of the SQV sets discussed above although not sufficiently reliable for use in predicting toxicity results at this site. It is possible that the quotient approach has merit, but it needs to be optimized on a site-specific basis. Both of the quotient methods tested here were developed based on data sets for marine and estuarine waters throughout the United States. The PEL-Q quotient method was specifically optimized for predicting acute amphipod toxicity in the data set used to develop the PEL-Q and therefore may not be optimal for the Portland Harbor data set, because it is clear that different chemicals are affecting different endpoints.

A.4 SUMMARY OF RELIABILITY RESULTS FOR EXISTING SQV SETS

None of the existing SQV sets perform well enough to use them in predicting biological effects at the Portland Harbor Superfund Site. The lower thresholds (the TELs, TECs, and LELs) are far too conservative to be useful because they classify all or nearly all stations as hits (low efficiency). The higher thresholds (the PECs, PELs, and SELs) are more successful at predicting toxic effects. None of the existing SQV sets perform well enough to use them in predicting biological effects at the Portland Harbor Superfund Site. The lower thresholds (the TELs, TECs, and LELs) are far too conservative to be useful because they classify all or nearly all stations as hits (low efficiency). The higher thresholds (the PECs, PELs, and SELs) are more successful at predicting toxic effects, yet the error rates are still high enough that substantial portions of the Study Area could be incorrectly classified as contributing to adverse effects.

Error rates are still high enough that substantial portions of the Study Area could be incorrectly classified as contributing to adverse effects. It is possible that the development of a site-specific SQV set or predictive model could reduce error rates.

A.5 REFERENCES

Avocet. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: Development and recommendation of SQVs for freshwater sediments in Washington State. Publication No. 03-09-088. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA.

Avocet, SAIC. 2002. Development of freshwater sediment quality values for use in Washington State. Final report. Publication No. 02-09-050. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA and Science Applications International Corporation (SAIC), Bothell, WA.

CCME. 2002. Canadian sediment quality guidelines for the protection of aquatic life: summary tables. Canadian Environmental Quality Guidelines, 1999, updated 2001, updated 2002 [online]. Canadian Council of Ministers of the Environment. Updated 2002. [Cited March 2005]. Available from: http://www.ccme.ca/assets/pdf/e7_002.pdf.

Fairey R, Long ER, Roberts CA, Anderson BS, Phillips BM, Hunt JW, Puckett HM, Wilson CJ. 2001. An evaluation of methods for calculating mean sediment quality guideline quotients as indicators of contamination and acute toxicity to amphipods by chemical mixtures. Environ Toxicol Chem 20(10):276-2286.

Ingersoll CG, MacDonald DD, Wang N, Crane JL, Field LJ, Haverland PS, Kemble NE, Lindskoog RA, Severn C, Smorong DE. 2000. Prediction of sediment toxicity using consensus-based freshwater sediment quality guidelines. US Environmental Protection Agency, Chicago, IL.

Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environ Toxicol Chem 17(4):714-727.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39:20-31.

Macfarlane MW, MacDonald DD, Ingersoll CG. 2002. Criteria for contaminated sites: criteria for managing contaminated sediment in British Columbia. Technical appendix. Ministry of Water, Land and Air Protection, Victoria, BC, Canada.

Persaud D, Jaagumagi R, Hayton A. 1993. Guidelines for the protection and management of aquatic sediment quality in Ontario. ISBN 0-7729-9248-7. Ontario Ministry of Environment and Energy, Toronto, ON.

APPENDIX B. CHEMICAL DATA SET FOR EXISTING SQV SETS

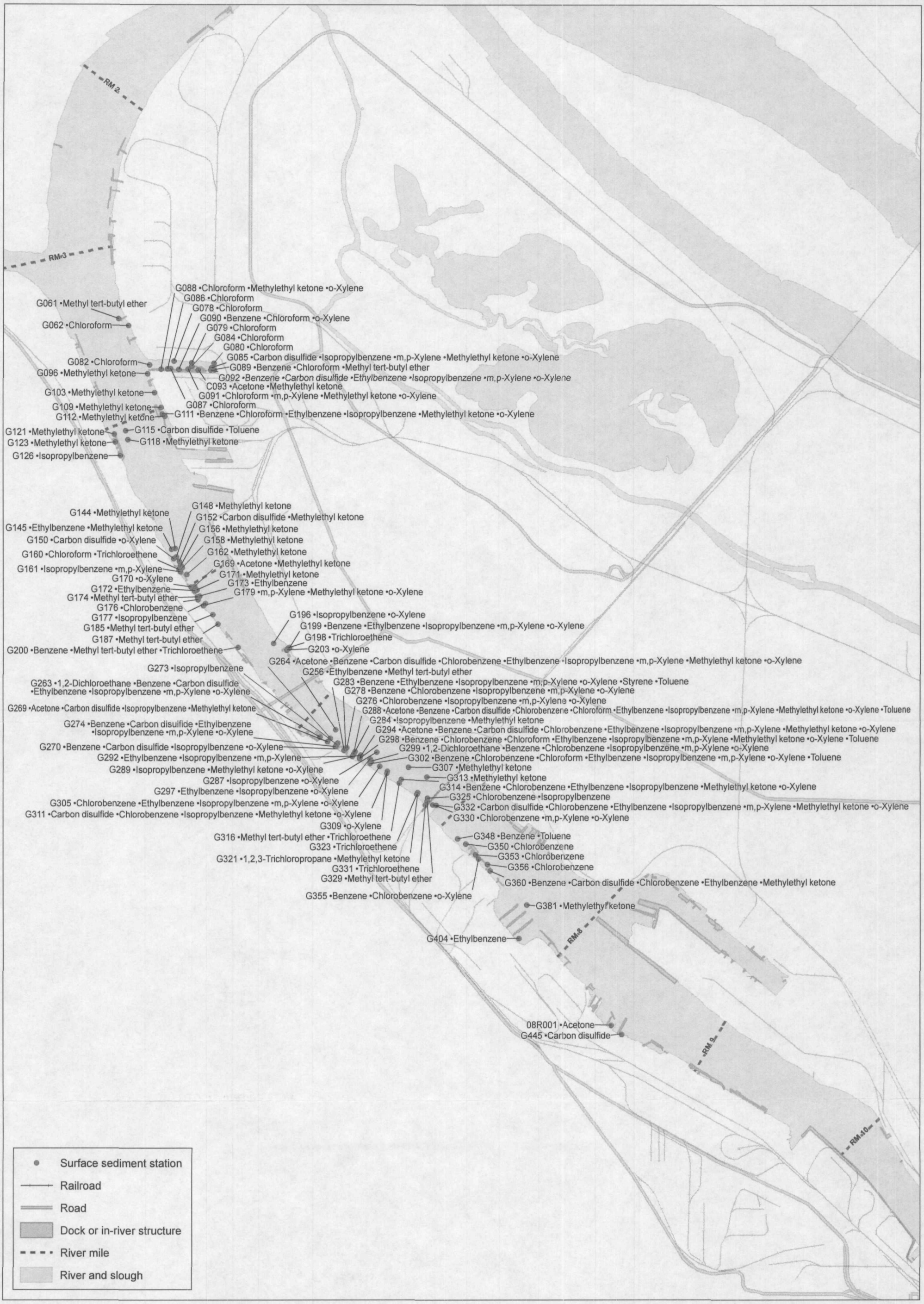


Figure B-1. Surface sediment stations with VOCs excluded from toxicity modeling

Wind Ward environmental LLC

Prepared by LSM 2-13-06 Map 2176

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\

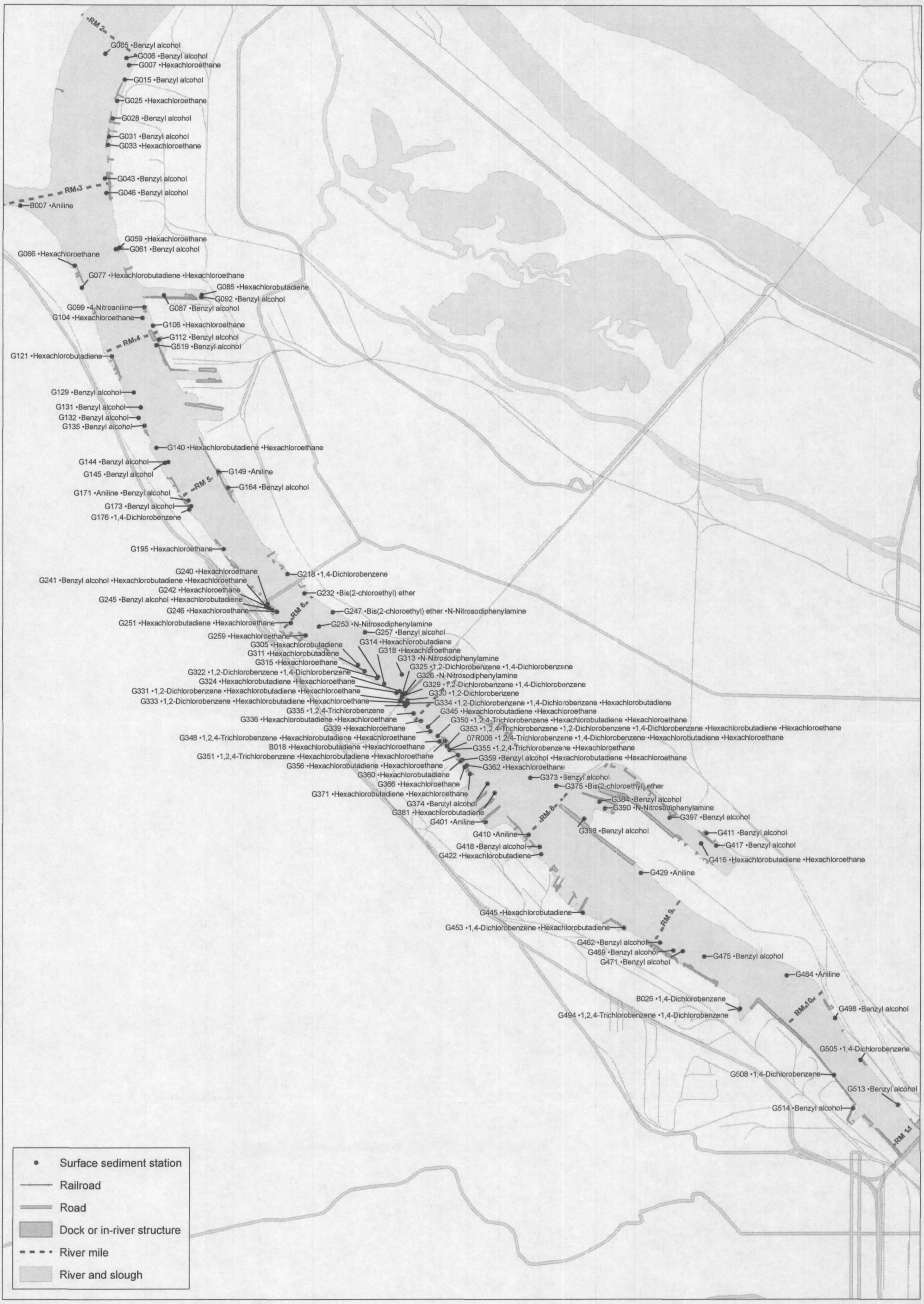


Figure B-2. Surface sediment stations with SVOCs excluded from toxicity modeling

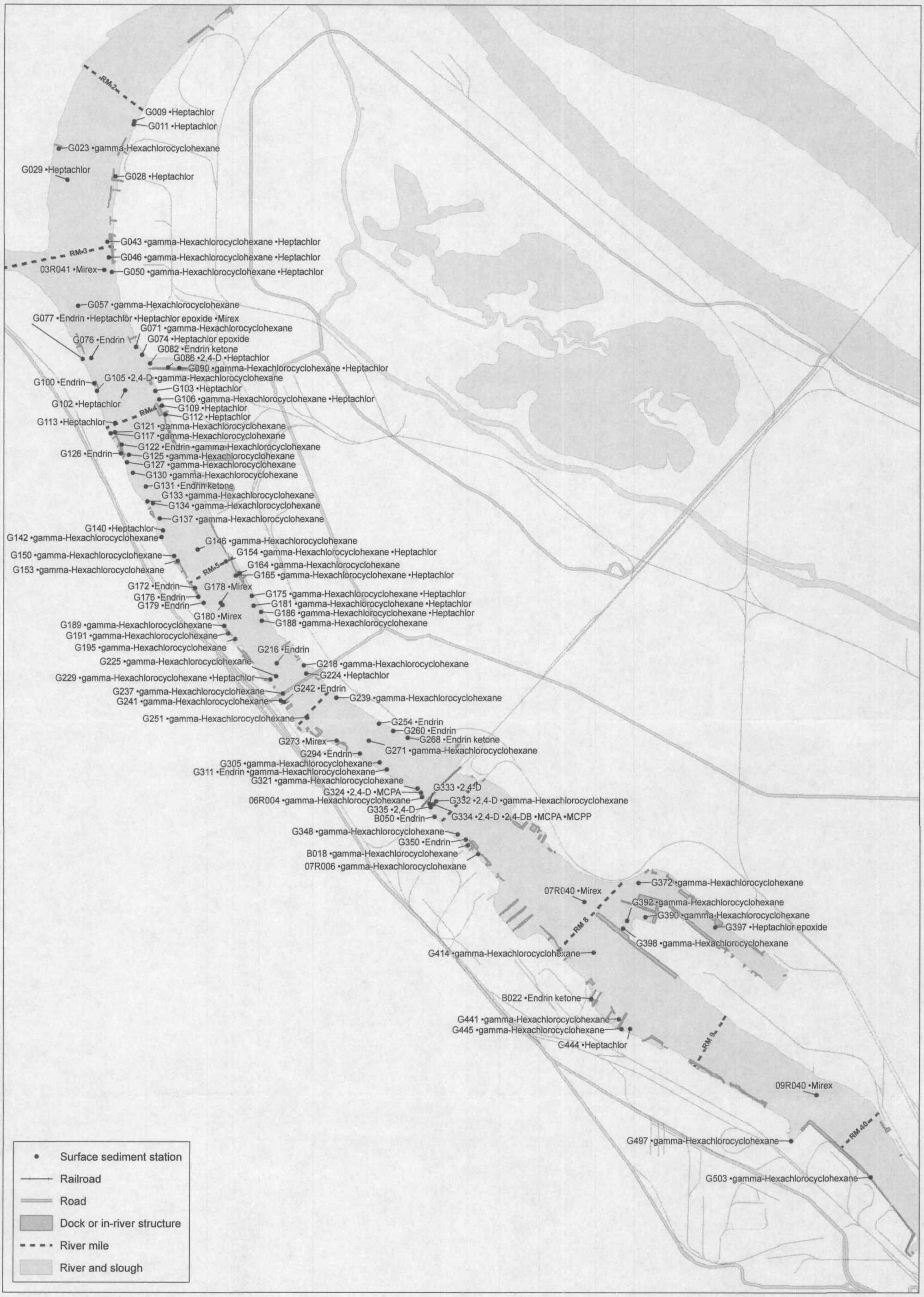


Figure B-3. Surface sediment stations with herbicides and pesticides excluded from toxicity modeling

0 0.25 0.5 1 1.5 Miles

0 0.5 1 2 Kilometers

N

Wind Ward environmental LLC

Prepared by LSM 2-9-06 Map 217

W:\Projects\01-25-01 Portland Harbor ERA\Data\

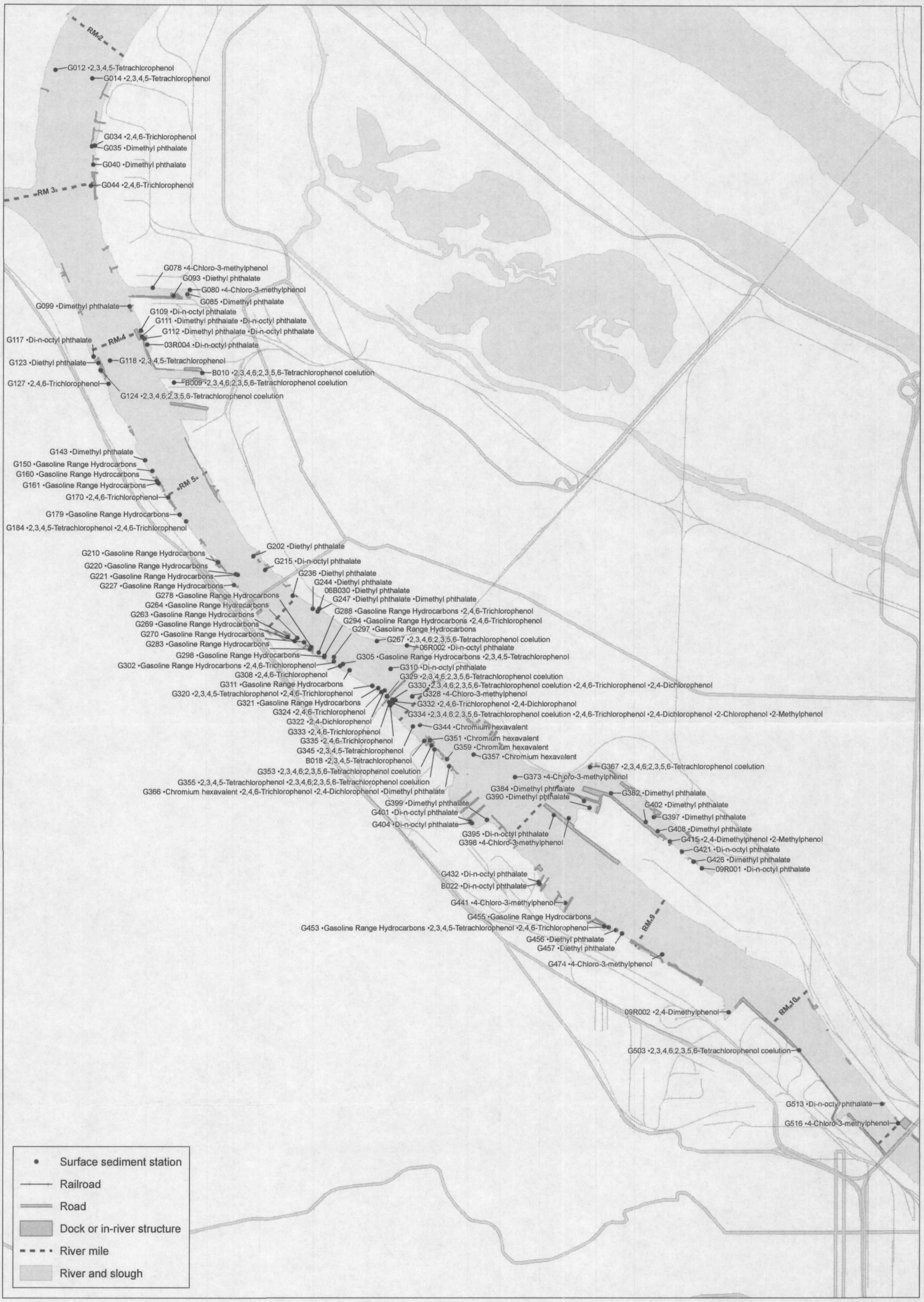


Figure B-4. Surface sediment stations with metal, petroleum, phenols and phthalates excluded from toxicity modeling

Wind Ward environmental LLC

Prepared by LSM 2-15-06 Map 2179

W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Bioassay\

APPENDIX C. RESULTS FOR MULTIVARIATE ANALYSES

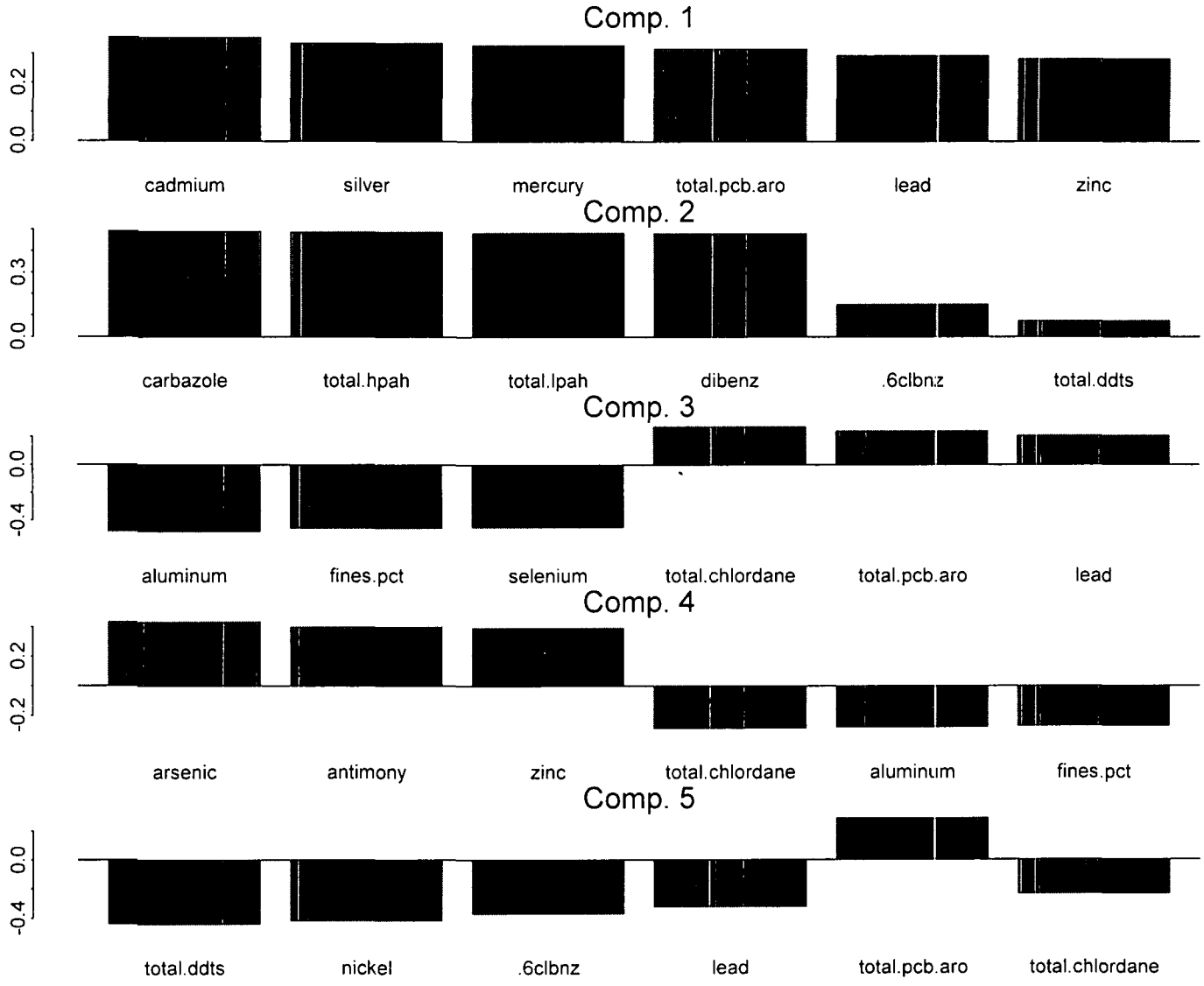


Figure C-1. Loadings plot, showing correlations between original (scaled) variables and first five principal components.

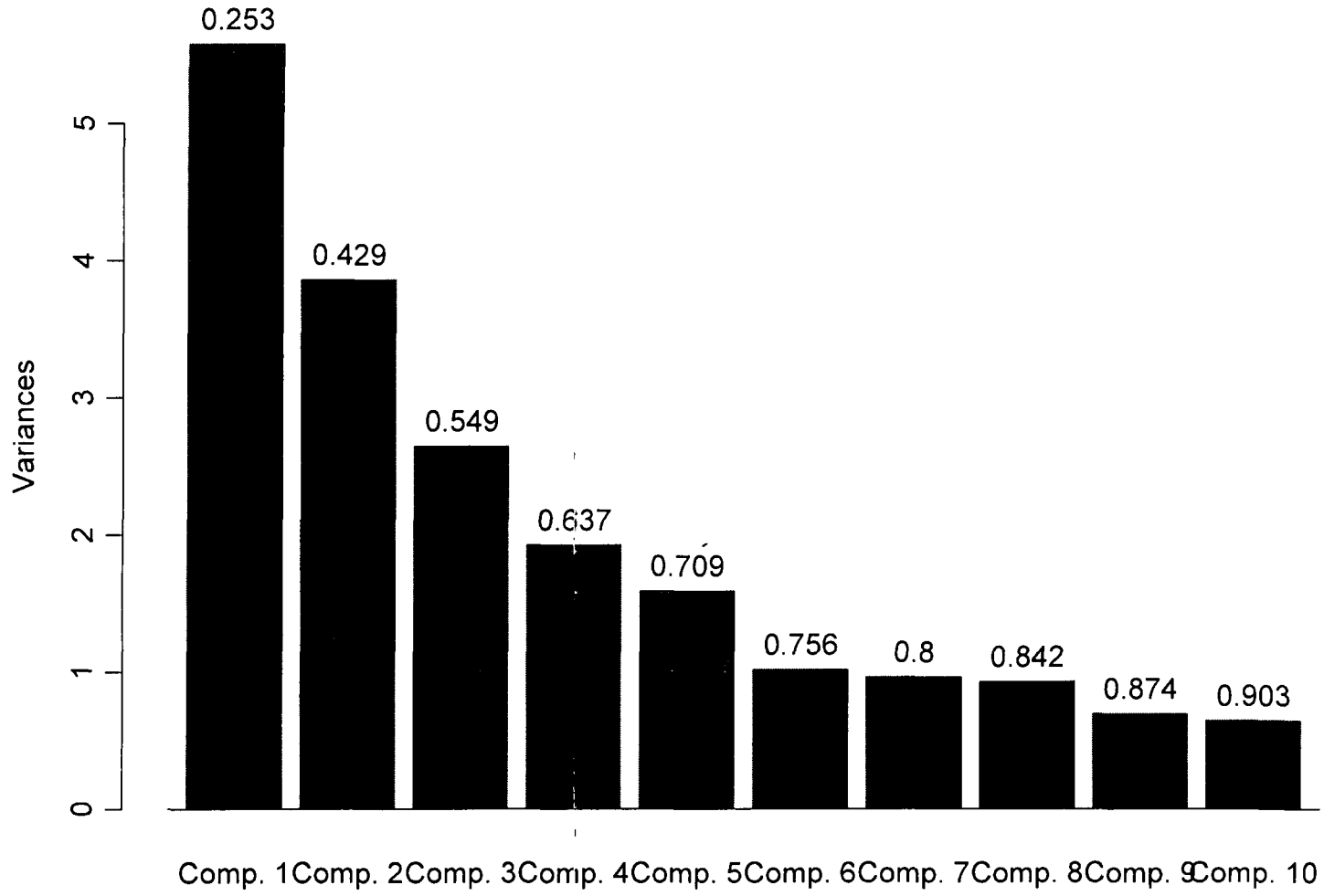


Figure C-2. Screeplot showing the cumulative variance explained by each successive principal component.

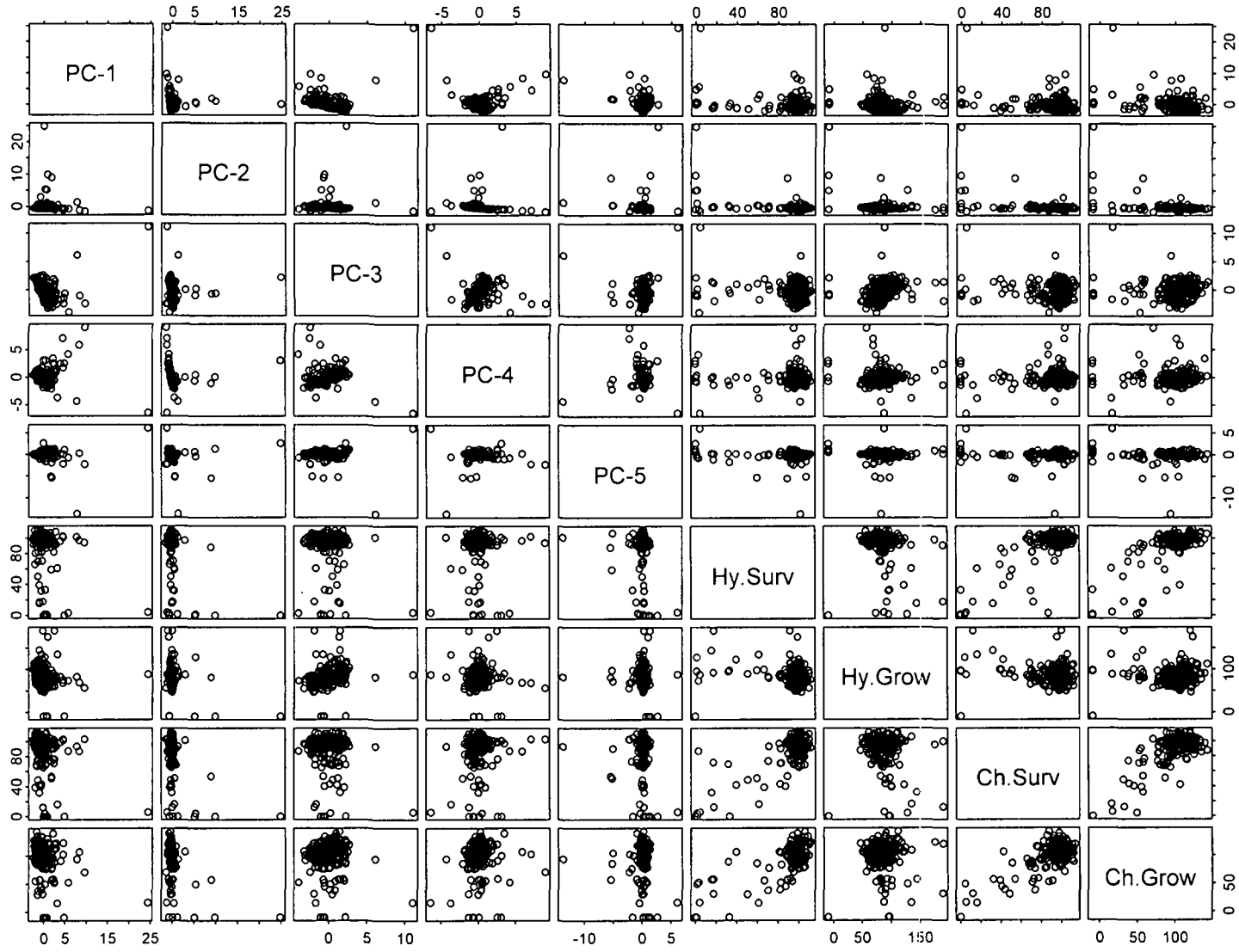


Figure C-3. Matrix of pair wise scatter plots between first five principal components and control-adjusted biological endpoints.

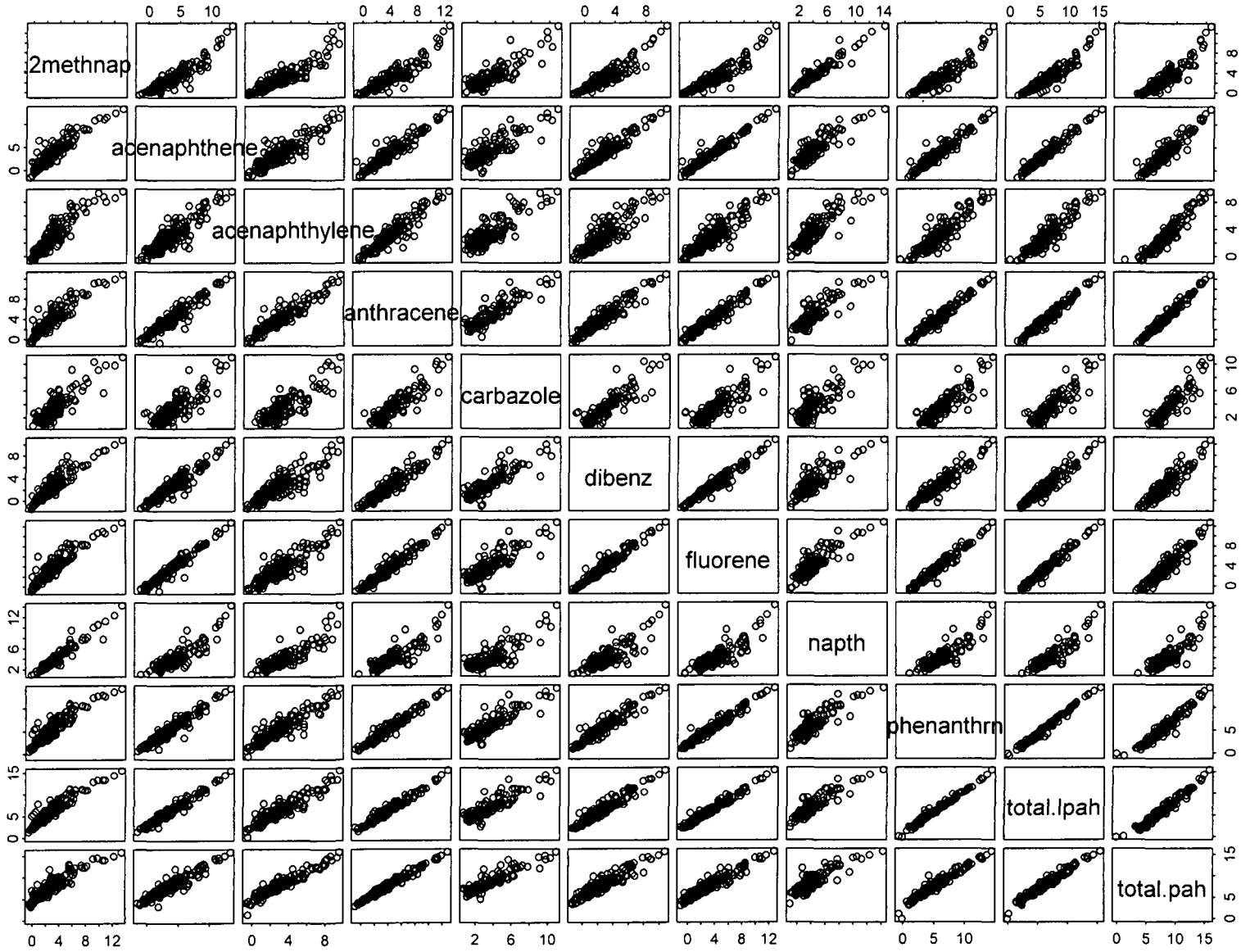


Figure C-4. Pairs plots for individual LPAHs, and sum of LPAHs and total PAHs, plus carbazole and dibenzofuran.

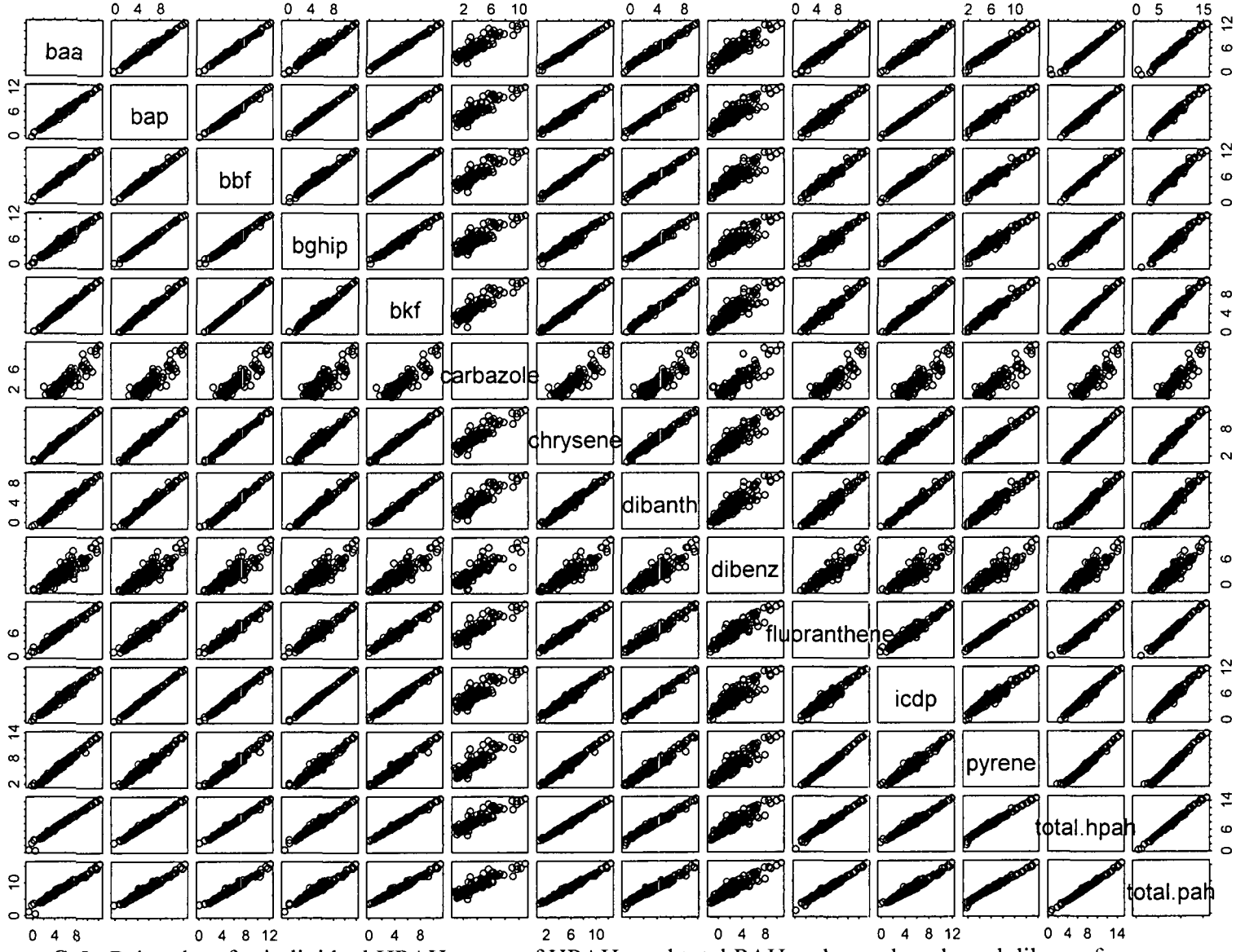


Figure C-5. Pairs plots for individual HPAHs, sum of HPAHs and total PAHs, plus carbazole and dibenzofuran.

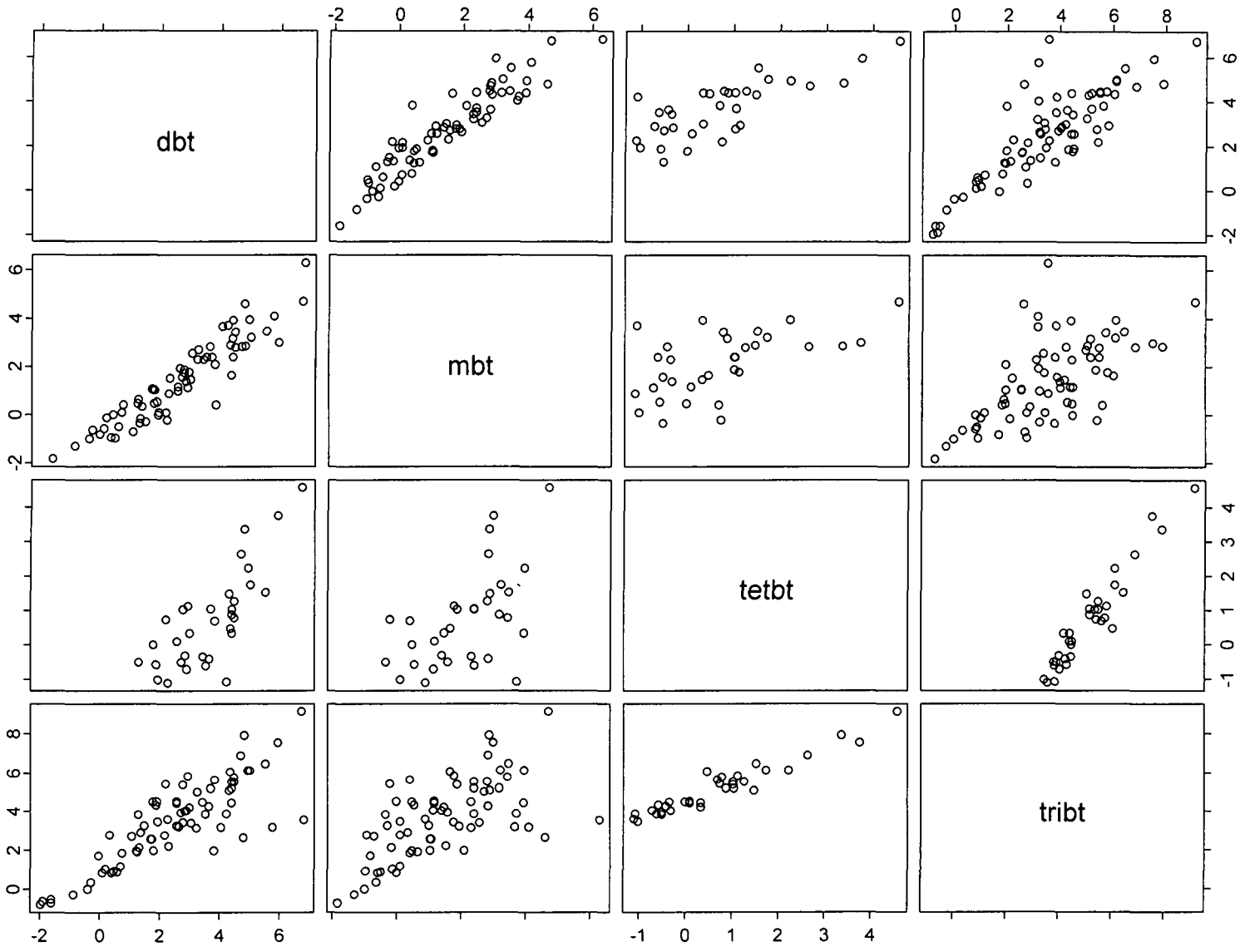


Figure C-6. Pairs plots for organotins.

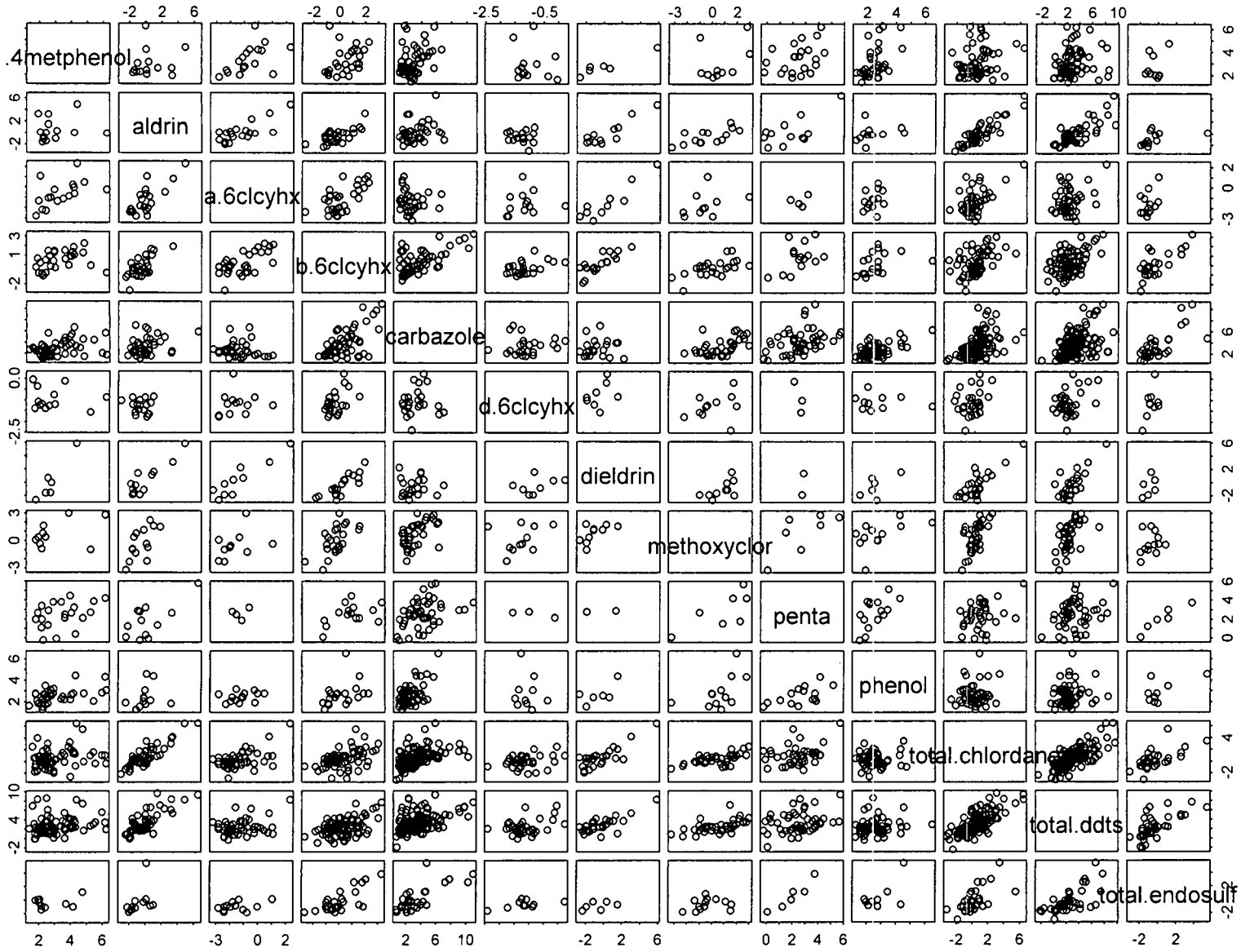


Figure C-7. Pairs plots for pesticides, phenols, and phthalates

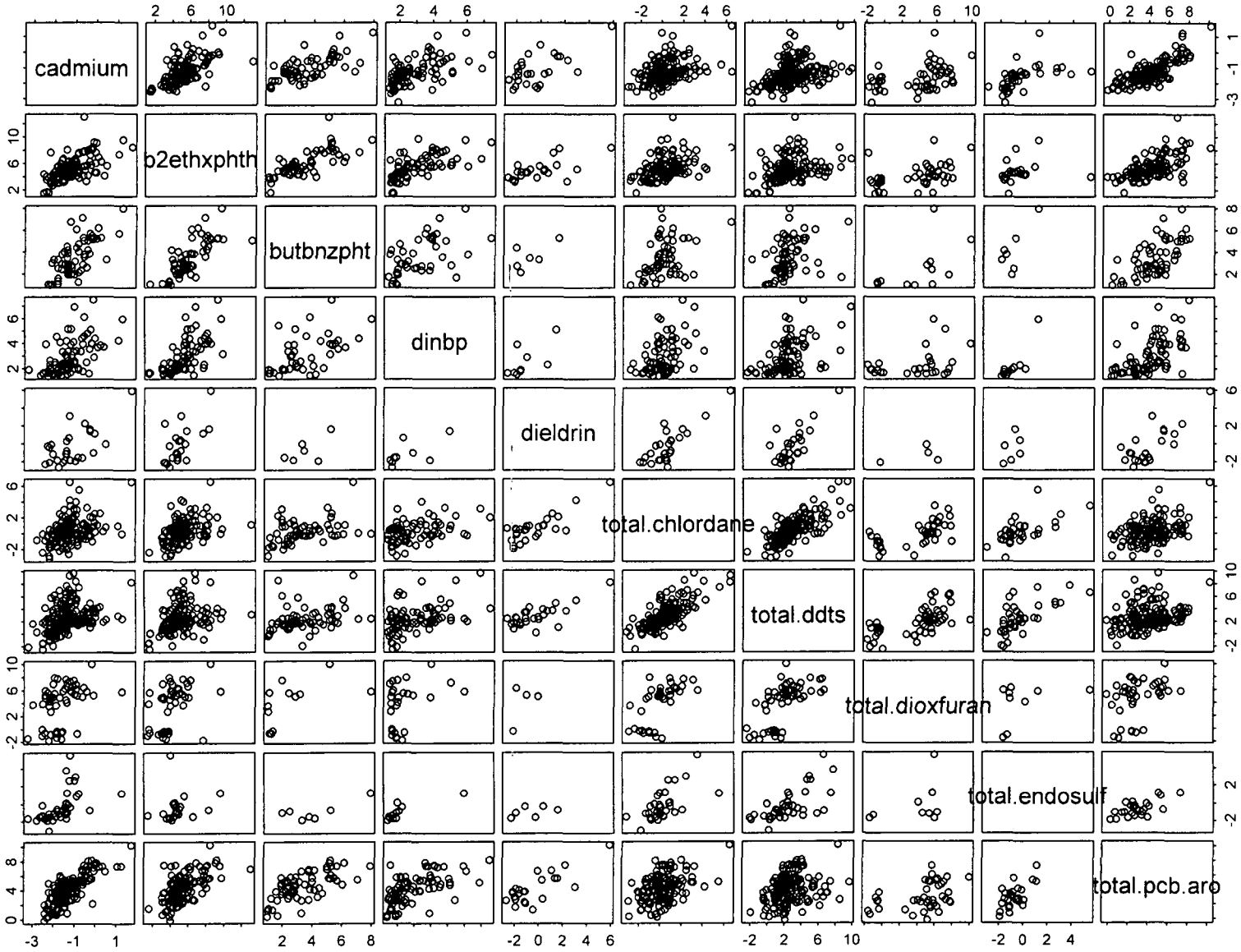


Figure C-8. Pairs plots for phthalates, cadmium, dieldrin, and totals for chlordane, ddts, dioxin/furans, endosulfan, and PCBs.

Table C-1. Description of stations in groups formed within a distance of 7.5.

GROUP #	STATIONS							
1	G283	G288	G294-1	G301				
2	G092	G455	G467	G473				
3	D1-1	G007-1	G009	G010	G011	G015	G017	G020
	G024	G026	G035	G066	G077	G085	G086	G088
	G090	G091	G093	G099	G103	G106	G109	G112
	G117	G121	G122	G123	G133	G136	G155	G166
	G172	G198	G199	G203-1	G228	G230	G240	G242
	G245	G267	G268	G276	G278	G282	G284	G296
	G302	G315	G334	G339	G345-1	G353-1	G359	G362-1
	G366	G368	G372-1	G376	G382	G384-1	G385	G387
	G392	G393	G398	G408	G415	G416	G426	G444
	G450-1	G457	G458	G461	G468	G469	G474	G477
	G497	U6TOC-2	U6TOC-3					
4	D2	G027	G033	G034	G038	G060	G062	G064
	G067	G073	G074	G078	G079	G080	G082	G083
	G089	G096	G105	G124	G127	G130	G139	G142
	G147	G157	G160	G161	G163	G164	G170	G176
	G178	G179	G180	G182	G184	G187	G197-1	G200
	G202	G204	G205	G206	G207	G209	G210	G212-1
	G213	G220	G221	G227	G231	G232	G234	G235
	G244	G247	G254	G260	G273	G274	G277	G280
	G292	G295	G303	G308	G316	G318	G320	G321
	G323	G324-1	G327	G329	G331	G333	G335	G336
	G342	G346	G347	G348	G350	G351	G352	G364
	G371	G377	G380	G386	G389	G396	G401	G403
	G405	G409	G413	G417	G420	G425	G430	G437
	G441	G454	G480	G492-1	U1C-1	U1C-2	U1C-3	U2C-1
	U2C-2	U2C-3	U3C-1	U3C-2	U3C-3	U4Q-1	U4Q-2	U4Q-3
	U5Q-1	U5Q-2	U5Q-3	U6TOC-1				
5	G019	G025	G383					
6	G111							
7	G263							
8	G264							
9	G270-1							
10	G298							
11	G311-1							
12	G355							
13	G360							
14	G367							
15	G390							
16	G445							
17	G453							
18	G456							

Table C-2a. Chemical and biological characteristics by cluster analysis groups defined in Table C-1: metals

	Group No.	Station Count	Mean Values By Group											
			Aluminum	Antimony	Arsenic	Cadmium	Chromium	Copper	Lead	Mercury	Nickel	Selenium	Silver	Zinc
	1	4	24,500	0.19	3.46	0.31	31	41	26	0.12	34	0.16	0.34	139
	2	4	16,825	6.97	8.61	1.49	67	139	185	0.19	36	0.14	0.30	464
	3	83	27,323	0.44	5.01	0.43	41	94	33	0.10	27	0.17	0.28	191
	4	124	17,584	0.25	3.72	0.19	24	38	20	0.08	21	0.07	0.15	105
	5	3	9,060	0.56	3.21	0.68	169	32	27	0.03	17	0.05	0.19	388
G111	6	1	21,600	1.94	15.50	3.51	103	216	120	0.27	78	0.06	0.62	1,940
G263	7	1	27,500	1.24	6.52	0.26	41	47	46	0.06	200	0.10	0.19	111
G264	8	1	24,400	0.46	4.94	0.37	34	55	27	0.17	52	0.20	0.53	160
G270-1	9	1	26,100	0.13	4.46	0.27	34	43	684	0.08	36	0.26	0.22	130
G298	10	1	20,200	0.19	2.23	0.25	26	31	18	0.08	26	0.09	0.36	101
G311-1	11	1	41,200	0.12	3.76	0.45	43	54	32	0.43	34	0.28	0.58	145
G355	12	1	17,300	1.93	8.37	0.29	48	147	1290	0.08	102	0.06	0.19	144
G360	13	1	20,800	0.26	7.43	0.35	58	101	33	0.06	29	0.11	0.26	136
G367	14	1	14,700	1.46	6.70	0.54	33	97	69	0.15	19	0.05	0.20	262
G390	15	1	28,000	1.89	16.50	0.66	51	1,080	102	0.30	32	0.20	0.64	731
G445	16	1	24,100	18.70	34.00	0.76	60	257	454	0.45	34	0.17	1.13	1,360
G453	17	1	20,200	6.37	7.30	5.41	146	120	956	2.01	22	0.16	4.44	561
G456	18	1	21,900	19.30	22.90	0.34	43	359	66	0.06	31	0.19	0.34	457

Table C-2b. Chemical and biological characteristics by cluster analysis groups defined in Table C-1: organic chemicals

			Mean Values By Group, cont.								
	Group #	Station Count	bis(2) Ethylhexyl Phthalate	Carbazole	Dibenz	Hexachloro-benzene	Total Chlordane	Total DDTs	Total HPAH	Total LPAH	Total PCBs
	1	4	395	10948	10350	4.1	4.1	590	1035475	893,725	25
	2	4	1103	188	160	1.0	7.7	48	9231	4220	1,258
	3	83	927	158	24	4.7	3.8	171	8212	1550	339
	4	124	140	53	75	3.4	1.4	51	14986	4226	48
	5	3	740	7	5	1.8	0.9	8	1138	139	666
G111	6	1	14,000	13	6	17.0	0.9	11	832	182	1,530
G263	7	1	52	220	42	0.1	0.3	34	39030	18,040	3.17
G264	8	1	940	30,000	2,600	1200.0	2.1	103	1312000	396,600	4.64
G270-1	9	1	230	760	420	2.3	4.7	39	164800	65,400	4.3
G298	10	1	340	56,000	46,000	0.3	0.6	2,309	2812000	5,134,000	14.2
G311-1	11	1	33	93	255	3.9	246.0	1,725	54825	27,230	170
G355	12	1	330	370	76	338.0	668.8	11,480	143000	6,338	-1000
G360	13	1	800	14	2.8	5.5	22.4	16,171	849	91	151
G367	14	1	440,000	31	12	34.0	2.8	21	4656	964	981
G390	15	1	3000	160	89	0.9	25.4	62	12650	3,078	1430
G445	16	1	310	52	67	2.2	1.7	135	3805	1620	271
G453	17	1	4500	29	190	4.6	659.8	3,928	2268	3,370	27,370
G456	18	1	460	110	86	0.2	9.0	28	4481	1,444	188.8

Table C-2c. Chemical and biological characteristics by cluster analysis groups defined in Table C-1: conventionals

			Mean Values By Group, cont.					
	Group #	Station Count	Fines (%)	Hyalella Survival	Hyalella Growth	Chironomus Survival	Chironomus Growth	Comments
	1	4	62	25	108	26	78	High PAHs, carbazole, dibenzofuran
	2	4	35	73	82	69	107	
	3	83	71	96	80	90	100	
	4	124	30	95	87	94	105	
	5	3	2	96	152	97	120	
G111	6	1	45	97	69	87	107	High As, Zn, and Cu
G263	7	1	47	107	72	90	104	High nickel
G264	8	1	69	89	81	53	57	High PAHs, carbazole, dibenzofuran, and hexachlorobenzene
G270-1	9	1	61	93	70	74	77	High lead
G298	10	1	56	0		0		Highest PAHs, carbazole, dibenzofuran; high DDTs; no survival
G311-1	11	1	89	61	135	16	16	High chlordane and aluminum
G355	12	1	19	101	83	93	94	High chlordane, DDTs and lead; but good survival
G360	13	1	63	59	98	51	87	Highest DDTs
G367	14	1	18	101	83	101	100	High b2ethxphth
G390	15	1	83	3	73	87	53	High As, Zn, and Cu
G445	16	1	71	94	57	103	71	High Sb, As, Zn, and Cu
G453	17	1	64	4	88	5	17	Highest PCBs; high chlordane and DDTs and low survival
G456	18	1	41	103	69	103	102	High Sb, As, Zn, and Cu

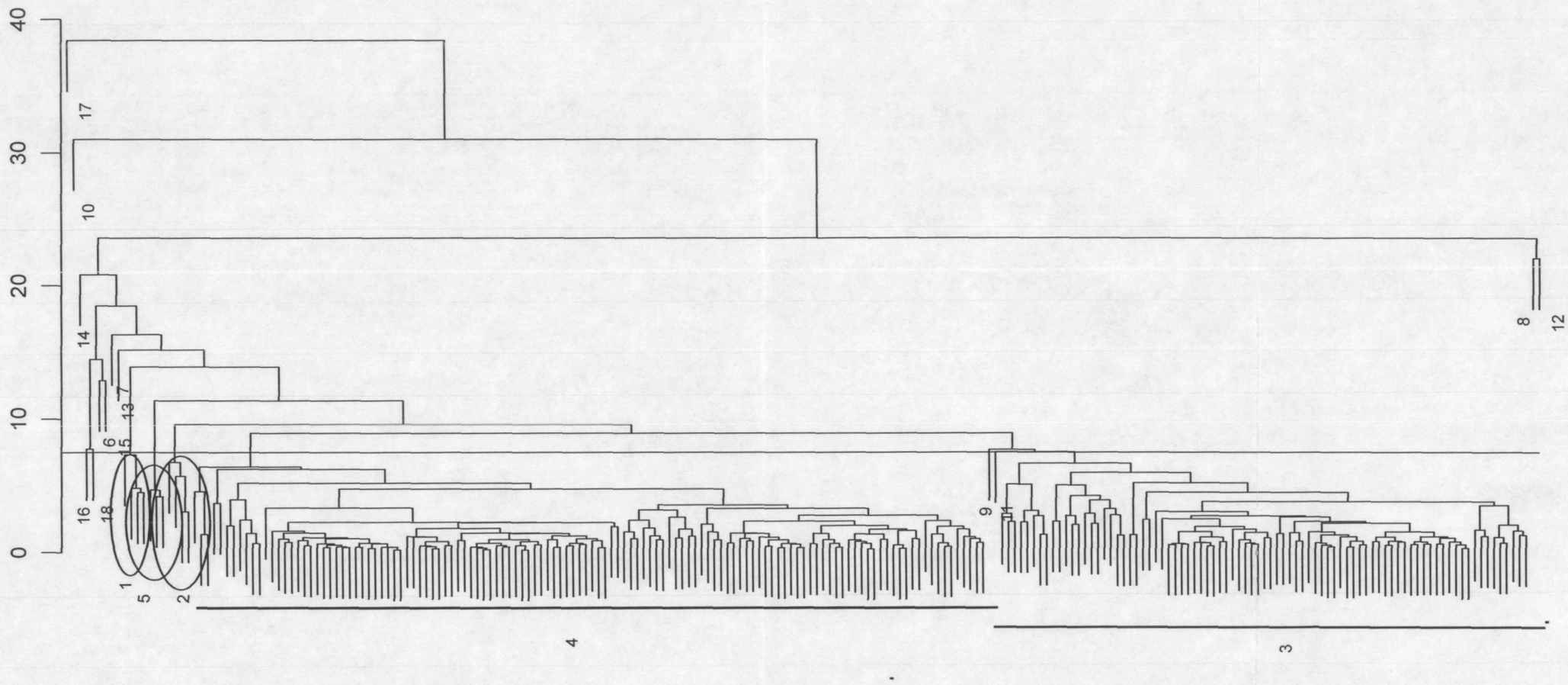


Figure C-9. Dendrogram of stations

APPENDIX D. FLOATING PERCENTILE MODEL DETAILS

This appendix can be found on the accompanying compact disk.

APPENDIX E. RESULTS OF INDIVIDUAL LOGISTIC REGRESSION MODELS

APPENDIX E. RESULTS OF INDIVIDUAL LOGISTIC REGRESSION MODELS

The results for the individual logistic regression models (LRMs) are presented in Table E-1.

- Columns 1 and 2 present the biological and chemical endpoints such that each row in this table represents an individual LRM.
- Columns 3 and 4 present the total number of samples and the number of toxic samples retained in the screened data set, respectively (i.e., after the low concentration toxic stations have been removed; Step 4, Section 5.3.1).
- Columns 5 and 6 present the Chi-square goodness-of-fit statistic and its p-value. Chi-square measures the change in deviance between the null model (intercept only) and the full model with slope and intercept (hence, a one-degree-of-freedom test). The measure of deviance is based on the log-likelihood, which indicates the probability of obtaining the observed toxicological responses, given the chemical responses and the specified model parameters. The log-likelihood function is maximized for the final slope and intercept parameters selected for the model, just as the sum of squares is minimized to select the slope and intercept in ordinary least squares regression. The Chi-square test for logistic regression is analogous to the F-test for ordinary least squares regression.
- Column 7 presents the R^2_L, the likelihood ratio R^2. It is equal to the change in deviance (the value of the Chi-square statistic) divided by the deviance associated with the null model. It is a substantive measure of the goodness-of-fit of the model that is not dependent upon sample size or the base rate of toxicity in the data (Menard 2000). It varies between zero and one, with zero indicating no relationship between chemistry and rate of toxicity and one indicating a perfect fit. R^2_L values of ##### indicates incalculable, when 0 toxic samples were retained in the screened data set.
- Columns 8 and 9 present the slope and intercept parameters, respectively, for the best fit model (B_0 and B_1 in Equation 1, Step 5, Section 5.3.1).
- Column 10 presents comments that indicate if any individual models were excluded (based on Chi-square p-values > 0.01 or fewer than two toxic stations retained in the screened data set) or were considered questionable or unreliable (based on $R^2_L < 0.20$ or fewer than five toxic stations retained in the screened data set).

A plot of the data and the best fit model for each of the models described in Table E-1 are shown in Figures E-1 to E-66. The nine models constructed for each chemical analyte are

shown on a single page. For each plot within a page, the $\log_{10}$ chemical concentration is shown on the x-axis, and the proportion of samples toxic within a concentration interval are shown on the y-axis. The symbol plotted at each (x,y) value is the number of samples within that concentration interval. All biological endpoints for an effects level are shown on a single row, and all endpoints for a species are shown in a single column. The title of each plot indicates the biological endpoint (e.g., hym.80 is Level 2 [80% difference] for *Hyalella* mortality).

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
Chironomus pooled									
chp.L3	Ammonia	220	19	30.69	0.00	0.24	6.00	-14.51	
chp.L3	Sulfide	191	15	53.81	0.00	0.51	3.80	-7.37	
chp.L3	Fines (%)	222	21	18.15	0.00	0.13	4.65	-10.32	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Aluminum	216	15	11.98	0.00	0.11	8.30	-38.96	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Antimony	144	7	16.20	0.00	0.29	2.63	-2.73	
chp.L3	Arsenic	214	13	14.81	0.00	0.15	4.30	-5.79	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Cadmium	214	15	18.63	0.00	0.17	3.31	-1.16	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Chromium	217	17	13.50	0.00	0.11	4.08	-8.84	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Copper	210	9	10.67	0.00	0.14	2.48	-7.67	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Lead	211	10	18.00	0.00	0.22	2.64	-7.19	
chp.L3	Mercury	212	15	27.38	0.00	0.25	4.05	1.15	
chp.L3	Nickel	206	16	15.29	0.00	0.14	6.06	-11.19	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Selenium	112	10	11.16	0.00	0.17	7.37	3.49	Questionable reliability ($R^2_L < 0.20$).
chp.L3	Silver	220	19	30.18	0.00	0.23	3.93	-0.03	
chp.L3	Zinc	210	9	17.26	0.00	0.23	4.29	-12.97	
chp.L3	Butyltin	65	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L3	Dibutyltin	68	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L3	Tributyltin	68	1	1.18	0.28	0.11	1.22	-6.69	Exclude (chi.p $\geq$ 0.01)
chp.L3	Acenaphthene	194	13	60.29	0.00	0.63	2.32	-9.02	
chp.L3	Anthracene	197	15	63.80	0.00	0.60	2.39	-8.97	
chp.L3	Fluorene	193	14	61.63	0.00	0.61	2.42	-8.71	
chp.L3	2-methylnaphthalene	193	12	55.24	0.00	0.61	2.40	-7.56	
chp.L3	Acenaphthylene	197	13	56.64	0.00	0.59	2.75	-8.55	
chp.L3	Naphthalene	159	14	55.06	0.00	0.58	2.61	-8.55	
chp.L3	Phenanthrene	205	14	62.79	0.00	0.61	2.36	-10.64	
chp.L3	Benzo(a)anthracene	206	16	59.41	0.00	0.53	2.38	-9.67	
chp.L3	Benzo(a)pyrene	204	14	60.53	0.00	0.59	2.70	-11.40	
chp.L3	Benzo(b)fluoranthene	204	15	58.59	0.00	0.55	2.59	-10.87	
chp.L3	Benzo(ghi)perylene	205	14	61.70	0.00	0.60	2.80	-11.52	
chp.L3	Benzo(k)fluoranthene	202	16	58.49	0.00	0.52	2.50	-9.18	
chp.L3	Chrysene	202	15	59.27	0.00	0.55	2.59	-10.79	
chp.L3	Dibenzanthracene	207	15	60.67	0.00	0.56	2.79	-9.02	
chp.L3	Fluoranthene	208	15	63.19	0.00	0.59	2.60	-11.72	
chp.L3	Indeno(c,d)pyrene	205	15	60.01	0.00	0.56	2.58	-10.52	
chp.L3	Pyrene	207	14	63.62	0.00	0.62	2.63	-12.14	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
chp.L3	Total LPAH	205	14	63.65	0.00	0.62	2.47	-11.71	
chp.L3	Total HPAH	210	15	61.84	0.00	0.57	2.56	-13.32	
chp.L3	Total PAHs	211	15	63.15	0.00	0.58	2.53	-13.51	
chp.L3	Diesel-range hydrocarbons	141	21	65.66	0.00	0.55	4.57	-13.32	
chp.L3	Residual organics	131	18	56.30	0.00	0.54	5.83	-19.41	
chp.L3	Dibenzofuran	194	15	57.97	0.00	0.55	2.53	-7.24	
chp.L3	Hexachlorobenzene	103	3	7.20	0.01	0.27	2.12	-4.30	Questionable reliability (fewer than 5 toxic stations retained)
chp.L3	Pentachlorodibenzofuran 12378	38	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
chp.L3	Pentachlorodibenzodioxin, homolo	47	2	3.88	0.05	0.23	2.59	-3.63	Exclude (chi.p ≥ 0.01)
chp.L3	TEQ mammal (0.5 detection limit)	56	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
chp.L3	Total dioxins/furans	56	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
chp.L3	Total PCBs	165	7	19.13	0.00	0.33	2.44	-9.08	
chp.L3	Aldrin	48	1	4.36	0.04	0.45	2.07	-5.83	Exclude (chi.p ≥ 0.01)
chp.L3	alpha-Hexachlorocyclohexane	48	1	9.71	0.00	1.00	22.87	-16.83	Exclude (only 1 hit retained)
chp.L3	beta-Hexachlorocyclohexane	83	11	25.37	0.00	0.39	4.09	-3.63	
chp.L3	delta-Hexachlorocyclohexane	33	1	3.69	0.05	0.41	9.97	-1.32	Exclude (chi.p ≥ 0.01)
chp.L3	Carbazole	155	16	53.23	0.00	0.52	2.65	-7.62	
chp.L3	Methoxychlor	35	1	1.09	0.30	0.12	2.34	-4.99	Exclude (chi.p ≥ 0.01)
chp.L3	cis-Nonachlor	53	5	10.77	0.00	0.33	4.68	-2.10	
chp.L3	trans-Nonachlor	72	2	2.72	0.10	0.15	2.80	-2.99	Exclude (chi.p ≥ 0.01)
chp.L3	Total chlordane	173	12	35.80	0.00	0.41	2.78	-4.24	
chp.L3	DDD	209	17	60.69	0.00	0.51	2.57	-6.45	
chp.L3	DDE	201	14	48.98	0.00	0.48	2.82	-5.66	
chp.L3	Total.ddt	176	5	22.65	0.00	0.50	1.96	-6.72	
chp.L3	Total.ddts	208	14	51.66	0.00	0.50	2.41	-7.27	
chp.L3	Total endosulfans	39	4	12.83	0.00	0.50	2.55	-3.17	Questionable reliability (fewer than 5 toxic stations retained)
chp.L3	4-Methylphenol	75	6	11.63	0.00	0.28	2.75	-6.99	
chp.L3	Pentachlorophenol	44	2	2.41	0.12	0.15	1.90	-5.79	Exclude (chi.p ≥ 0.01)
chp.L3	Phenol	66	1	10.35	0.00	1.00	16.15	-39.44	Exclude (only 1 hit retained)
chp.L3	bis(2-ethylhexyl) phthalate	141	1	2.26	0.13	0.19	1.34	-8.67	Exclude (chi.p ≥ 0.01)
chp.L3	Butylbenzyl phthalate	66	2	2.45	0.12	0.14	1.53	-6.38	Exclude (chi.p ≥ 0.01)
chp.L3	Dibutyl phthalate	94	6	12.93	0.00	0.29	2.36	-6.53	
chp.L2	Phenol	63	1	10.26	0.00	1.00	16.27	-39.75	Exclude (only 1 hit retained)
chp.L2	alpha-Hexachlorocyclohexane	46	1	9.62	0.00	1.00	23.00	-16.93	Exclude (only 1 hit retained)
chp.L2	Total LPAH	198	17	76.95	0.00	0.66	2.77	-12.43	
chp.L2	Phenanthrene	198	17	76.61	0.00	0.66	2.69	-11.36	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
chp.L2	Fluorene	186	17	74.63	0.00	0.66	2.70	-9.02	
chp.L2	Anthracene	190	18	76.90	0.00	0.65	2.69	-9.43	
chp.L2	2-methylnaphthalene	185	14	63.83	0.00	0.64	2.72	-7.85	
chp.L2	Acenaphthene	189	18	74.95	0.00	0.63	2.39	-8.36	
chp.L2	Benzo(a)pyrene	196	16	69.67	0.00	0.63	2.91	-11.85	
chp.L2	Acenaphthylene	189	15	65.06	0.00	0.62	2.93	-8.65	
chp.L2	Total PAHs	204	18	74.26	0.00	0.61	2.73	-13.99	
chp.L2	Pyrene	201	18	73.91	0.00	0.61	2.63	-11.47	
chp.L2	Dibenzanthracene	199	17	70.16	0.00	0.60	3.06	-9.44	
chp.L2	Dibenzofuran	187	18	70.57	0.00	0.60	2.89	-7.61	
chp.L2	Fluoranthene	202	19	74.10	0.00	0.59	2.68	-11.44	
chp.L2	Benzo(ghi)perylene	199	18	69.89	0.00	0.58	2.68	-10.46	
chp.L2	Indeno(c,d)pyrene	198	18	69.25	0.00	0.57	2.69	-10.44	
chp.L2	Total HPAH	204	19	71.58	0.00	0.57	2.59	-12.89	
chp.L2	Diesel-range hydrocarbons	139	26	75.68	0.00	0.56	5.31	-14.63	
chp.L2	Naphthalene	155	19	64.86	0.00	0.56	2.77	-8.15	
chp.L2	Benzo(k)fluoranthene	194	18	67.22	0.00	0.56	2.72	-9.57	
chp.L2	Benzo(b)fluoranthene	197	18	67.10	0.00	0.56	2.68	-10.77	
chp.L2	Dibutyl phthalate	94	9	32.92	0.00	0.55	4.60	-10.17	
chp.L2	Benzo(a)anthracene	199	19	69.12	0.00	0.55	2.54	-9.80	
chp.L2	Carbazole	150	19	62.79	0.00	0.55	3.07	-8.16	
chp.L2	Chrysene	196	19	68.50	0.00	0.55	2.63	-10.39	
chp.L2	DDD	203	19	63.63	0.00	0.50	2.49	-6.02	
chp.L2	Residual organics	127	22	58.59	0.00	0.50	5.83	-18.87	
chp.L2	Total endosulfans	37	4	12.66	0.00	0.50	2.50	-3.07	Questionable reliability (fewer than 5 toxic stations retained)
chp.L2	Total.ddt	168	5	22.45	0.00	0.50	1.93	-6.58	
chp.L2	Total.ddts	202	16	54.44	0.00	0.49	2.32	-6.77	
chp.L2	DDE	196	17	53.04	0.00	0.46	2.75	-5.15	
chp.L2	Sulfide	193	26	67.74	0.00	0.44	3.56	-5.91	
chp.L2	Aldrin	45	1	4.25	0.04	0.44	2.05	-5.78	Exclude (chi.p $\geq$ 0.01)
chp.L2	Total chlordane	166	12	35.85	0.00	0.42	2.75	-4.16	
chp.L2	delta-Hexachlorocyclohexane	31	1	3.57	0.06	0.40	9.89	-1.32	Exclude (chi.p $\geq$ 0.01)
chp.L2	Antimony	139	9	23.25	0.00	0.35	3.00	-2.43	
chp.L2	beta-Hexachlorocyclohexane	80	13	24.18	0.00	0.34	3.63	-3.05	
chp.L2	Lead	206	15	35.80	0.00	0.33	3.65	-8.40	
chp.L2	Total PCBs	162	11	26.24	0.00	0.33	2.39	-8.31	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
chp.L2	cis-Nonachlor	53	9	14.11	0.00	0.29	4.21	-1.16	
chp.L2	Mercury	209	22	40.03	0.00	0.28	4.48	2.04	
chp.L2	4-Methylphenol	74	9	15.19	0.00	0.28	2.75	-6.38	
chp.L2	Ammonia	217	26	43.33	0.00	0.27	6.56	-15.26	
chp.L2	bis(2-ethylhexyl) phthalate	136	2	5.62	0.02	0.27	1.56	-8.65	Exclude (chi.p $\geq$ 0.01)
chp.L2	Hexachlorobenzene	96	3	7.03	0.01	0.26	2.08	-4.21	Questionable reliability (fewer than 5 toxic stations retained)
chp.L2	Silver	216	25	37.27	0.00	0.24	4.07	0.44	
chp.L2	Zinc	205	14	24.02	0.00	0.24	4.41	-12.68	
chp.L2	Pentachlorodibenzo-p-dioxin homologs	47	2	3.88	0.05	0.23	2.59	-3.63	Exclude (chi.p $\geq$ 0.01)
chp.L2	Cadmium	208	19	26.66	0.00	0.21	3.79	-0.69	
chp.L2	Arsenic	209	18	20.03	0.00	0.16	4.54	-5.57	Questionable reliability (R^2_L < 0.20).
chp.L2	Selenium	110	15	13.97	0.00	0.16	6.72	3.53	Questionable reliability (R^2_L < 0.20).
chp.L2	Copper	202	11	13.13	0.00	0.15	2.53	-7.50	Questionable reliability (R^2_L < 0.20).
chp.L2	Butylbenzyl phthalate	65	3	3.62	0.06	0.15	1.55	-5.95	Exclude (chi.p $\geq$ 0.01)
chp.L2	Pentachlorophenol	40	2	2.31	0.13	0.15	1.82	-5.56	Exclude (chi.p $\geq$ 0.01)
chp.L2	Fines (%)	221	30	24.27	0.00	0.14	4.31	-9.25	Questionable reliability (R^2_L < 0.20).
chp.L2	Nickel	204	24	18.78	0.00	0.13	6.33	-11.06	Questionable reliability (R^2_L < 0.20).
chp.L2	Aluminum	214	23	18.39	0.00	0.13	8.53	-39.45	Questionable reliability (R^2_L < 0.20).
chp.L2	Methoxychlor	34	1	1.06	0.30	0.12	2.29	-4.93	Exclude (chi.p $\geq$ 0.01)
chp.L2	trans-Nonachlor	70	4	3.57	0.06	0.12	2.27	-2.21	Exclude (chi.p $\geq$ 0.01)
chp.L2	Tributyltin	65	1	1.16	0.28	0.11	1.19	-6.58	Exclude (chi.p $\geq$ 0.01)
chp.L2	Chromium	213	23	16.32	0.00	0.11	4.09	-8.47	Questionable reliability (R^2_L < 0.20).
chp.L2	Dibutyltin	65	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L2	Butyltin	62	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L2	TEQ mammal (0.5 detection limit)	56	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L2	Total dioxins/furans	56	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L2	1,2,3,7,8-Pentachlorodibenzofuran	38	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
chp.L1	Ammonia	200	39	48.10	0.00	0.24	5.54	-12.36	
chp.L1	Sulfide	165	27	62.91	0.00	0.43	3.35	-5.41	
chp.L1	Fines (%)	202	41	38.06	0.00	0.19	5.06	-10.05	Questionable reliability (R^2_L < 0.20).
chp.L1	Aluminum	194	33	28.96	0.00	0.16	9.59	-43.53	Questionable reliability (R^2_L < 0.20).
chp.L1	Antimony	116	10	22.73	0.00	0.33	2.84	-2.11	
chp.L1	Arsenic	183	22	23.08	0.00	0.17	4.83	-5.36	Questionable reliability (R^2_L < 0.20).
chp.L1	Cadmium	183	23	37.02	0.00	0.27	4.84	0.16	
chp.L1	Chromium	191	31	19.22	0.00	0.11	4.31	-8.29	Questionable reliability (R^2_L < 0.20).
chp.L1	Copper	174	13	18.35	0.00	0.20	2.81	-7.75	Questionable reliability (R^2_L < 0.20).

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
chp.L1	Lead	178	17	36.91	0.00	0.33	3.62	-8.00	
chp.L1	Mercury	185	28	42.42	0.00	0.27	4.19	2.31	
chp.L1	Nickel	182	29	20.90	0.00	0.13	7.07	-11.72	Questionable reliability ($R^2_L < 0.20$).
chp.L1	Selenium	102	21	25.64	0.00	0.25	8.89	5.81	
chp.L1	Silver	189	28	40.70	0.00	0.26	4.17	0.82	
chp.L1	Zinc	179	18	30.38	0.00	0.26	4.99	-13.54	
chp.L1	Butyltin	51	5	6.60	0.01	0.20	2.17	-4.31	Exclude (chi.p ≥ 0.01)
chp.L1	Dibutyltin	52	3	9.57	0.00	0.42	3.54	-9.36	Questionable reliability (fewer than 5 toxic stations retained)
chp.L1	Tributyltin	51	3	11.54	0.00	0.51	3.26	-10.93	Questionable reliability (fewer than 5 toxic stations retained)
chp.L1	Acenaphthene	158	16	68.15	0.00	0.66	2.48	-8.88	
chp.L1	Anthracene	161	18	71.67	0.00	0.64	2.61	-9.08	
chp.L1	Fluorene	156	16	68.40	0.00	0.66	2.73	-9.26	
chp.L1	2-methylnaphthalene	155	13	58.35	0.00	0.65	2.64	-7.71	
chp.L1	Acenaphthylene	161	16	62.15	0.00	0.60	2.74	-7.87	
chp.L1	Naphthalene	131	18	61.16	0.00	0.58	2.79	-8.19	
chp.L1	Phenanthrene	168	17	71.71	0.00	0.65	2.60	-10.96	
chp.L1	Benzo(a)anthracene	167	17	64.29	0.00	0.59	2.63	-10.27	
chp.L1	Benzo(a)pyrene	168	17	65.84	0.00	0.60	2.69	-10.77	
chp.L1	Benzo(b)fluoranthene	168	18	63.23	0.00	0.55	2.62	-10.42	
chp.L1	Benzo(ghi)perylene	169	18	67.21	0.00	0.59	2.67	-10.33	
chp.L1	Benzo(k)fluoranthene	167	19	63.19	0.00	0.53	2.56	-8.84	
chp.L1	Chrysene	168	19	64.31	0.00	0.54	2.55	-10.01	
chp.L1	Dibenzanthracene	171	18	65.14	0.00	0.57	2.79	-8.48	
chp.L1	Fluoranthene	172	19	70.48	0.00	0.59	2.65	-11.25	
chp.L1	Indeno(c,d)pyrene	170	19	66.06	0.00	0.55	2.56	-9.75	
chp.L1	Pyrene	172	18	70.26	0.00	0.61	2.59	-11.20	
chp.L1	Total LPAH	168	17	71.94	0.00	0.65	2.68	-11.98	
chp.L1	Total HPAH	174	19	67.90	0.00	0.57	2.55	-12.59	
chp.L1	Total PAHs	174	18	70.05	0.00	0.61	2.67	-13.57	
chp.L1	Diesel-range hydrocarbons	126	27	70.12	0.00	0.54	4.86	-13.25	
chp.L1	Residual organics	113	23	53.02	0.00	0.46	5.28	-17.00	
chp.L1	Dibenzofuran	158	17	64.97	0.00	0.60	2.91	-7.72	
chp.L1	Hexachlorobenzene	77	2	5.16	0.02	0.28	1.99	-4.53	Exclude (chi.p ≥ 0.01)
chp.L1	1,2,3,7,8-Pentachlorodibenzofuran	33	0	0.00	0.98	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
chp.L1	Pentachlorodibenzo-p-dioxin homologs	42	2	3.91	0.05	0.24	2.50	-3.44	Exclude (chi.p ≥ 0.01)
chp.L1	TEQ mammal (0.5 detection limit)	48	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
chp.L1	Total dioxins/furans	48	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
chp.L1	Total PCBs	141	14	29.43	0.00	0.32	2.24	-7.40	
chp.L1	Aldrin	31	1	3.72	0.05	0.42	1.88	-5.32	Exclude (chi.p ≥ 0.01)
chp.L1	alpha-Hexachlorocyclohexane	33	2	5.59	0.02	0.37	3.19	-2.19	Exclude (chi.p ≥ 0.01)
chp.L1	beta-Hexachlorocyclohexane	64	16	24.09	0.00	0.33	3.77	-2.39	
chp.L1	delta-Hexachlorocyclohexane	26	3	1.84	0.17	0.10	3.35	-0.82	Exclude (chi.p ≥ 0.01)
chp.L1	Carbazole	129	20	59.76	0.00	0.54	3.08	-8.00	
chp.L1	Methoxychlor	27	2	3.61	0.06	0.25	4.54	-6.25	Exclude (chi.p ≥ 0.01)
chp.L1	cis-Nonachlor	40	8	12.21	0.00	0.30	4.68	-1.21	
chp.L1	trans-Nonachlor	56	6	6.90	0.01	0.18	2.98	-1.36	Questionable reliability (R^2_L < 0.20).
chp.L1	Total chlordane	140	11	34.72	0.00	0.45	2.93	-4.20	
chp.L1	DDD	173	18	60.61	0.00	0.52	2.51	-6.04	
chp.L1	DDE	167	17	52.72	0.00	0.48	2.82	-5.05	
chp.L1	Total.ddt	140	6	23.82	0.00	0.48	1.85	-6.02	
chp.L1	Total.ddts	171	14	51.09	0.00	0.53	2.39	-7.06	
chp.L1	Total endosulfans	30	4	12.09	0.00	0.51	2.32	-2.60	Questionable reliability (fewer than 5 toxic stations retained)
chp.L1	4-Methylphenol	65	12	26.33	0.00	0.42	4.29	-8.22	
chp.L1	Pentachlorophenol	38	3	3.61	0.06	0.17	1.91	-5.20	Exclude (chi.p ≥ 0.01)
chp.L1	Phenol	53	2	10.63	0.00	0.62	5.72	-12.72	Questionable reliability (fewer than 5 toxic stations retained)
chp.L1	bis(2-ethylhexyl) phthalate	114	2	5.95	0.01	0.30	1.60	-8.46	Exclude (chi.p ≥ 0.01)
chp.L1	Butylbenzyl phthalate	53	3	4.78	0.03	0.21	1.96	-6.45	Exclude (chi.p ≥ 0.01)
chp.L1	Dibutyl phthalate	79	10	35.64	0.00	0.59	5.75	-11.81	
Hyalella mortality									
hym.L3	Ammonia	225	10	17.32	0.00	0.21	5.87	-15.03	
hym.L3	Sulfide	198	8	29.71	0.00	0.44	2.99	-7.19	
hym.L3	Fines (%)	229	14	13.14	0.00	0.12	4.95	-11.34	Questionable reliability (R^2_L < 0.20).
hym.L3	Aluminum	223	8	10.15	0.00	0.15	11.07	-51.95	Questionable reliability (R^2_L < 0.20).
hym.L3	Antimony	155	4	9.58	0.00	0.26	2.48	-3.43	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	Arsenic	221	6	7.89	0.00	0.14	4.21	-6.64	Questionable reliability (R^2_L < 0.20).
hym.L3	Cadmium	222	9	12.84	0.00	0.17	3.26	-1.82	Questionable reliability (R^2_L < 0.20).
hym.L3	Chromium	224	10	11.19	0.00	0.14	4.42	-10.05	Questionable reliability (R^2_L < 0.20).
hym.L3	Copper	219	4	9.34	0.00	0.23	3.36	-10.48	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	Lead	220	5	8.95	0.00	0.19	2.38	-7.58	Questionable reliability (R^2_L < 0.20).
hym.L3	Mercury	220	9	14.99	0.00	0.20	3.44	0.03	Questionable reliability (R^2_L < 0.20).
hym.L3	Nickel	213	10	8.06	0.00	0.10	4.71	-9.78	Questionable reliability (R^2_L < 0.20).
hym.L3	Selenium	117	6	7.09	0.01	0.15	7.27	2.81	Questionable reliability (R^2_L < 0.20).

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L3	Silver	228	13	20.90	0.00	0.21	3.59	-0.69	
hym.L3	Zinc	220	5	11.96	0.00	0.25	4.43	-14.06	
hym.L3	Butyltin	68	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Dibutyltin	71	0	0.00	0.96	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Tributyltin	70	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Acenaphthene	203	8	38.80	0.00	0.58	2.01	-9.00	
hym.L3	Anthracene	206	10	43.83	0.00	0.55	2.07	-8.83	
hym.L3	Fluorene	202	9	40.12	0.00	0.55	2.04	-8.46	
hym.L3	2-methylnaphthalene	203	8	35.19	0.00	0.52	1.77	-6.95	
hym.L3	Acenaphthylene	207	9	39.19	0.00	0.53	2.47	-8.54	
hym.L3	Naphthalene	166	9	37.02	0.00	0.53	2.05	-8.01	
hym.L3	Phenanthrene	215	10	42.87	0.00	0.53	1.95	-9.72	
hym.L3	Benzo(a)anthracene	213	9	41.43	0.00	0.56	2.49	-11.23	
hym.L3	Benzo(a)pyrene	213	9	42.61	0.00	0.57	2.63	-12.06	
hym.L3	Benzo(b)fluoranthene	212	9	41.21	0.00	0.55	2.64	-12.12	
hym.L3	Benzo(ghi)perylene	214	9	43.79	0.00	0.59	2.78	-12.45	
hym.L3	Benzo(k)fluoranthene	209	9	41.53	0.00	0.56	2.67	-10.93	
hym.L3	Chrysene	210	9	41.39	0.00	0.56	2.58	-11.82	
hym.L3	Dibenzanthracene	216	10	39.83	0.00	0.49	2.42	-8.76	
hym.L3	Fluoranthene	217	10	44.25	0.00	0.55	2.37	-11.66	
hym.L3	Indeno(c,d)pyrene	213	9	43.07	0.00	0.58	2.73	-12.18	
hym.L3	Pyrene	216	9	44.55	0.00	0.60	2.50	-12.58	
hym.L3	Total LPAH	215	10	42.94	0.00	0.53	1.98	-10.33	
hym.L3	Total HPAH	218	9	43.61	0.00	0.58	2.61	-14.70	
hym.L3	Total PAHs	219	9	43.88	0.00	0.58	2.48	-14.38	
hym.L3	Diesel-range hydrocarbons	141	13	46.02	0.00	0.53	3.43	-11.45	
hym.L3	Residual organics	132	11	40.61	0.00	0.54	4.74	-17.17	
hym.L3	Dibenzofuran	203	10	37.13	0.00	0.47	2.02	-6.83	
hym.L3	Hexachlorobenzene	113	3	7.45	0.01	0.27	2.18	-4.40	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	1,2,3,7,8-Pentachlorodibenzofuran	39	0	0.00	0.97	0.00	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Pentachlorodibenzo-p-dioxin homologs	47	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	TEQ mammal (0.5 detection limit)	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Total dioxins/furans	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Total PCBs	170	3	15.74	0.00	0.52	4.00	-15.13	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	Aldrin	51	1	4.47	0.03	0.45	2.09	-5.89	Exclude (chi.p ≥ 0.01)
hym.L3	alpha-Hexachlorocyclohexane	50	1	9.79	0.00	1.00	22.63	-16.65	Exclude (only 1 hit retained)

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L3	beta-Hexachlorocyclohexane	88	7	19.72	0.00	0.40	4.26	-4.55	
hym.L3	delta-Hexachlorocyclohexane	36	0	0.00	0.98	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Carbazole	156	7	36.28	0.00	0.64	2.62	-9.55	
hym.L3	Methoxychlor	39	1	1.27	0.26	0.14	2.46	-5.11	Exclude (chi.p ≥ 0.01)
hym.L3	cis-Nonachlor	57	4	10.05	0.00	0.35	4.88	-2.48	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	trans-Nonachlor	74	0	0.00	0.96	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L3	Total chlordane	181	9	28.98	0.00	0.40	2.59	-4.60	
hym.L3	DDD	218	12	52.18	0.00	0.56	2.86	-7.81	
hym.L3	DDE	209	10	43.00	0.00	0.54	3.01	-6.63	
hym.L3	Total.ddt	187	4	18.69	0.00	0.48	1.93	-7.03	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	Total.ddts	219	11	45.44	0.00	0.52	2.50	-7.99	
hym.L3	Total endosulfans	41	1	3.75	0.05	0.40	2.08	-5.06	Exclude (chi.p ≥ 0.01)
hym.L3	4-Methylphenol	76	4	7.39	0.01	0.24	2.52	-7.08	Questionable reliability (fewer than 5 toxic stations retained)
hym.L3	Pentachlorophenol	46	2	2.53	0.11	0.15	1.94	-5.87	Exclude (chi.p ≥ 0.01)
hym.L3	Phenol	66	1	10.35	0.00	1.00	16.15	-39.44	Exclude (only 1 hit retained)
hym.L3	bis(2-ethylhexyl) phthalate	146	1	2.24	0.13	0.19	1.35	-8.75	Exclude (chi.p ≥ 0.01)
hym.L3	Butylbenzyl phthalate	66	2	2.45	0.12	0.14	1.53	-6.38	Exclude (chi.p ≥ 0.01)
hym.L3	Dibutyl phthalate	94	5	12.07	0.00	0.31	2.45	-6.97	
hym.L2	Ammonia	224	11	21.57	0.00	0.25	6.55	-16.39	
hym.L2	Sulfide	198	10	37.77	0.00	0.48	3.27	-7.31	
hym.L2	Fines (%)	228	15	13.53	0.00	0.12	4.73	-10.86	Questionable reliability (R^2_L < 0.20).
hym.L2	Aluminum	221	8	10.03	0.00	0.15	11.02	-51.73	Questionable reliability (R^2_L < 0.20).
hym.L2	Antimony	154	4	9.54	0.00	0.26	2.48	-3.43	Questionable reliability (fewer than 5 toxic stations retained)
hym.L2	Arsenic	219	6	7.84	0.01	0.14	4.20	-6.62	Questionable reliability (R^2_L < 0.20).
hym.L2	Cadmium	220	9	12.74	0.00	0.17	3.25	-1.82	Questionable reliability (R^2_L < 0.20).
hym.L2	Chromium	222	10	11.08	0.00	0.14	4.40	-10.01	Questionable reliability (R^2_L < 0.20).
hym.L2	Copper	218	5	9.08	0.00	0.19	2.97	-9.37	Questionable reliability (R^2_L < 0.20).
hym.L2	Lead	218	5	8.91	0.00	0.19	2.37	-7.56	Questionable reliability (R^2_L < 0.20).
hym.L2	Mercury	219	10	14.72	0.00	0.18	3.25	0.01	Questionable reliability (R^2_L < 0.20).
hym.L2	Nickel	212	11	8.06	0.00	0.09	4.60	-9.51	Questionable reliability (R^2_L < 0.20).
hym.L2	Selenium	115	6	6.91	0.01	0.15	7.21	2.77	Questionable reliability (R^2_L < 0.20).
hym.L2	Silver	227	14	22.24	0.00	0.21	3.63	-0.58	
hym.L2	Zinc	218	5	11.90	0.00	0.25	4.42	-14.03	
hym.L2	Butyltin	68	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Dibutyltin	71	0	0.00	0.96	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Tributyltin	70	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L2	Acenaphthene	202	9	41.63	0.00	0.57	1.97	-8.61	
hym.L2	Anthracene	205	11	46.24	0.00	0.54	2.04	-8.53	
hym.L2	Fluorene	201	10	43.09	0.00	0.54	2.03	-8.22	
hym.L2	2-methylnaphthalene	202	9	36.85	0.00	0.50	1.75	-6.64	
hym.L2	Acenaphthylene	206	10	42.50	0.00	0.53	2.46	-8.32	
hym.L2	Naphthalene	165	10	38.84	0.00	0.51	2.05	-7.76	
hym.L2	Phenanthrene	214	11	46.48	0.00	0.54	1.98	-9.65	
hym.L2	Benzo(a)anthracene	212	10	44.87	0.00	0.56	2.48	-11.01	
hym.L2	Benzo(a)pyrene	212	10	46.17	0.00	0.57	2.61	-11.80	
hym.L2	Benzo(b)fluoranthene	211	10	44.76	0.00	0.56	2.64	-11.90	
hym.L2	Benzo(ghi)perylene	213	10	47.44	0.00	0.59	2.76	-12.13	
hym.L2	Benzo(k)fluoranthene	208	10	45.00	0.00	0.56	2.66	-10.69	
hym.L2	Chrysene	209	10	44.80	0.00	0.56	2.57	-11.59	
hym.L2	Dibenzanthracene	215	11	43.45	0.00	0.50	2.45	-8.68	
hym.L2	Fluoranthene	216	11	47.95	0.00	0.55	2.40	-11.58	
hym.L2	Indeno(c,d)pyrene	212	10	46.78	0.00	0.58	2.72	-11.92	
hym.L2	Pyrene	215	10	47.93	0.00	0.59	2.47	-12.23	
hym.L2	Total LPAH	214	11	46.22	0.00	0.53	2.00	-10.23	
hym.L2	Total HPAH	217	10	47.12	0.00	0.58	2.59	-14.39	
hym.L2	Total PAHs	218	10	47.23	0.00	0.58	2.47	-14.08	
hym.L2	Diesel-range hydrocarbons	141	15	47.72	0.00	0.50	3.38	-10.96	
hym.L2	Residual organics	132	13	38.17	0.00	0.45	4.14	-14.85	
hym.L2	Dibenzofuran	202	11	40.10	0.00	0.47	2.05	-6.74	
hym.L2	Hexachlorobenzene	112	3	7.45	0.01	0.27	2.17	-4.39	Questionable reliability (fewer than 5 toxic stations retained)
hym.L2	1,2,3,7,8-Pentachlorodibenzofuran	39	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Pentachlorodibenzo-p-dioxin homologs	47	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	TEQ mammal (0.5 detection limit)	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Total dioxins/furans	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Total PCBs	169	3	15.71	0.00	0.52	4.00	-15.11	Questionable reliability (fewer than 5 toxic stations retained)
hym.L2	Aldrin	50	1	4.43	0.04	0.45	2.08	-5.88	Exclude (chi.p ≥ 0.01)
hym.L2	alpha-Hexachlorocyclohexane	50	2	5.41	0.02	0.32	3.17	-2.80	Exclude (chi.p ≥ 0.01)
hym.L2	beta-Hexachlorocyclohexane	87	8	19.56	0.00	0.37	3.86	-4.05	
hym.L2	delta-Hexachlorocyclohexane	36	0	0.00	0.98	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Carbazole	156	8	38.67	0.00	0.61	2.57	-9.07	
hym.L2	Methoxychlor	39	1	1.27	0.26	0.14	2.46	-5.11	Exclude (chi.p ≥ 0.01)
hym.L2	cis-Nonachlor	57	4	10.05	0.00	0.35	4.88	-2.48	Questionable reliability (fewer than 5 toxic stations retained)

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L2	trans-Nonachlor	74	0	0.00	0.96	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L2	Total chlordane	180	9	29.01	0.00	0.41	2.58	-4.58	
hym.L2	DDD	216	12	52.00	0.00	0.56	2.85	-7.79	
hym.L2	DDE	208	10	42.93	0.00	0.54	3.00	-6.62	
hym.L2	DDT	186	4	18.67	0.00	0.48	1.93	-7.01	Questionable reliability (fewer than 5 toxic stations retained)
hym.L2	Total DDTs	217	11	45.28	0.00	0.52	2.49	-7.96	
hym.L2	Total endosulfans	41	1	3.75	0.05	0.40	2.08	-5.06	Exclude (chi.p ≥ 0.01)
hym.L2	4-Methylphenol	76	5	8.25	0.00	0.22	2.41	-6.59	
hym.L2	Pentachlorophenol	45	2	2.52	0.11	0.15	1.92	-5.81	Exclude (chi.p ≥ 0.01)
hym.L2	Phenol	66	1	10.35	0.00	1.00	16.15	-39.44	Exclude (only 1 hit retained)
hym.L2	bis(2-ethylhexyl) phthalate	145	1	2.23	0.13	0.19	1.35	-8.73	Exclude (chi.p ≥ 0.01)
hym.L2	Butylbenzyl phthalate	66	2	2.45	0.12	0.14	1.53	-6.38	Exclude (chi.p ≥ 0.01)
hym.L2	Dibutyl phthalate	94	5	12.07	0.00	0.31	2.45	-6.97	
hym.L1	Ammonia	213	13	21.06	0.00	0.22	5.76	-14.44	
hym.L1	Sulfide	189	14	47.67	0.00	0.48	3.69	-7.21	
hym.L1	Fines (%)	222	22	19.93	0.00	0.14	4.72	-10.37	Questionable reliability (R^2_L < 0.20).
hym.L1	Aluminum	215	15	13.19	0.00	0.12	8.72	-40.79	Questionable reliability (R^2_L < 0.20).
hym.L1	Antimony	144	5	9.56	0.00	0.22	2.25	-3.06	
hym.L1	Arsenic	208	8	11.34	0.00	0.17	4.50	-6.52	Questionable reliability (R^2_L < 0.20).
hym.L1	Cadmium	209	10	14.73	0.00	0.18	3.40	-1.63	Questionable reliability (R^2_L < 0.20).
hym.L1	Chromium	211	12	14.16	0.00	0.15	4.66	-10.19	Questionable reliability (R^2_L < 0.20).
hym.L1	Copper	207	7	9.87	0.00	0.16	2.67	-8.32	Questionable reliability (R^2_L < 0.20).
hym.L1	Lead	205	5	9.82	0.00	0.21	2.49	-7.74	
hym.L1	Mercury	210	14	20.14	0.00	0.20	3.42	0.58	Questionable reliability (R^2_L < 0.20).
hym.L1	Nickel	203	14	9.30	0.00	0.09	4.69	-9.31	Questionable reliability (R^2_L < 0.20).
hym.L1	Selenium	107	7	8.58	0.00	0.17	7.76	3.38	Questionable reliability (R^2_L < 0.20).
hym.L1	Silver	214	14	22.93	0.00	0.22	3.73	-0.49	
hym.L1	Zinc	206	6	14.54	0.00	0.27	4.61	-14.28	
hym.L1	Butyltin	66	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L1	Dibutyltin	66	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L1	Tributyltin	66	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hym.L1	Acenaphthene	192	10	46.07	0.00	0.59	2.12	-8.76	
hym.L1	Anthracene	195	13	53.38	0.00	0.56	2.25	-8.71	
hym.L1	Fluorene	192	11	48.35	0.00	0.57	2.23	-8.53	
hym.L1	2-methylnaphthalene	190	9	37.94	0.00	0.52	1.87	-6.78	
hym.L1	Acenaphthylene	194	10	43.46	0.00	0.55	2.59	-8.53	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L1	Naphthalene	157	11	41.35	0.00	0.52	2.24	-7.93	
hym.L1	Phenanthrene	202	12	51.19	0.00	0.56	2.17	-10.07	
hym.L1	Benzo(a)anthracene	201	12	47.39	0.00	0.52	2.40	-10.18	
hym.L1	Benzo(a)pyrene	200	10	49.24	0.00	0.62	3.02	-13.26	
hym.L1	Benzo(b)fluoranthene	200	11	47.12	0.00	0.55	2.70	-11.80	
hym.L1	Benzo(ghi)perylene	201	11	49.07	0.00	0.58	2.75	-11.72	
hym.L1	Benzo(k)fluoranthene	199	12	47.22	0.00	0.52	2.54	-9.74	
hym.L1	Chrysene	201	13	47.50	0.00	0.49	2.36	-10.12	
hym.L1	Dibenzanthracene	205	12	47.71	0.00	0.52	2.65	-8.94	
hym.L1	Fluoranthene	205	13	51.34	0.00	0.53	2.38	-11.03	
hym.L1	Indeno(c,d)pyrene	202	12	47.85	0.00	0.53	2.49	-10.52	
hym.L1	Pyrene	205	12	51.03	0.00	0.56	2.38	-11.28	
hym.L1	Total LPAH	202	12	50.99	0.00	0.56	2.20	-10.78	
hym.L1	Total HPAH	207	13	49.25	0.00	0.51	2.32	-12.36	
hym.L1	Total PAHs	207	12	50.54	0.00	0.55	2.42	-13.29	
hym.L1	Diesel-range hydrocarbons	139	17	51.76	0.00	0.50	3.78	-11.65	
hym.L1	Residual organics	128	14	40.44	0.00	0.46	4.62	-16.09	
hym.L1	Dibenzofuran	193	12	45.30	0.00	0.50	2.26	-6.96	
hym.L1	Hexachlorobenzene	107	3	7.28	0.01	0.27	2.14	-4.34	Questionable reliability (fewer than 5 toxic stations retained)
hym.L1	1,2,3,7,8-Pentachlorodibenzofuran	39	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
hym.L1	Pentachlorodibenzo-p-dioxin homologs	47	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
hym.L1	TEQ mammal (0.5 detection limit)	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
hym.L1	Total dioxins/furans	58	0	0.00	0.97	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
hym.L1	Total PCBs	164	4	19.27	0.00	0.51	4.07	-14.88	Questionable reliability (fewer than 5 toxic stations retained)
hym.L1	Aldrin	48	1	4.36	0.04	0.45	2.08	-5.86	Exclude (chi.p $\geq$ 0.01)
hym.L1	alpha-Hexachlorocyclohexane	49	2	5.36	0.02	0.32	3.14	-2.79	Exclude (chi.p $\geq$ 0.01)
hym.L1	beta-Hexachlorocyclohexane	83	8	19.47	0.00	0.37	3.86	-3.98	
hym.L1	delta-Hexachlorocyclohexane	33	0	0.00	0.98	######	0.00	-11.20	Exclude (chi.p $\geq$ 0.01)
hym.L1	Carbazole	153	13	42.11	0.00	0.47	2.40	-7.33	
hym.L1	Methoxychlor	37	1	1.20	0.27	0.13	2.37	-5.02	Exclude (chi.p $\geq$ 0.01)
hym.L1	cis-Nonachlor	56	5	11.22	0.00	0.33	4.81	-2.13	
hym.L1	trans-Nonachlor	72	1	0.75	0.39	0.07	1.96	-3.72	Exclude (chi.p $\geq$ 0.01)
hym.L1	Total chlordane	171	8	27.54	0.00	0.43	2.60	-4.76	
hym.L1	DDD	205	11	52.17	0.00	0.61	3.18	-8.73	
hym.L1	DDE	199	10	42.22	0.00	0.53	2.97	-6.54	
hym.L1	DDT	178	4	18.36	0.00	0.48	1.91	-6.97	Questionable reliability (fewer than 5 toxic stations retained)

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hym.L1	Total DDTs	207	11	44.40	0.00	0.52	2.47	-7.88	
hym.L1	Total endosulfans	37	1	3.70	0.05	0.40	2.02	-4.87	Exclude (chi.p ≥ 0.01)
hym.L1	4-Methylphenol	76	5	8.25	0.00	0.22	2.41	-6.59	
hym.L1	Pentachlorophenol	44	2	2.44	0.12	0.15	1.90	-5.78	Exclude (chi.p ≥ 0.01)
hym.L1	Phenol	64	1	10.29	0.00	1.00	16.26	-39.71	Exclude (only 1 hit retained)
hym.L1	bis(2-ethylhexyl) phthalate	143	1	2.25	0.13	0.19	1.34	-8.70	Exclude (chi.p ≥ 0.01)
hym.L1	Butylbenzyl phthalate	66	2	2.45	0.12	0.14	1.53	-6.38	Exclude (chi.p ≥ 0.01)
hym.L1	Dibutyl phthalate	92	6	12.71	0.00	0.29	2.34	-6.50	
Hyalella pooled									
hyp.L3	Ammonia	205	36	39.14	0.00	0.21	4.77	-10.98	
hyp.L3	Sulfide	158	13	34.76	0.00	0.39	2.62	-5.65	
hyp.L3	Fines (%)	225	56	57.69	0.00	0.23	5.85	-11.16	
hyp.L3	Aluminum	218	49	48.93	0.00	0.21	11.24	-50.35	
hyp.L3	Antimony	129	10	25.83	0.00	0.37	3.17	-2.27	
hyp.L3	Arsenic	194	25	28.05	0.00	0.19	4.94	-5.38	Questionable reliability (R^2_L < 0.20).
hyp.L3	Cadmium	194	26	30.41	0.00	0.20	3.66	-0.24	Questionable reliability (R^2_L < 0.20).
hyp.L3	Chromium	206	38	21.21	0.00	0.11	4.07	-7.76	Questionable reliability (R^2_L < 0.20).
hyp.L3	Copper	186	17	40.69	0.00	0.36	4.13	-10.19	
hyp.L3	Lead	186	17	29.86	0.00	0.26	3.08	-7.18	
hyp.L3	Mercury	189	24	37.06	0.00	0.26	4.09	1.91	
hyp.L3	Nickel	198	41	28.87	0.00	0.14	7.69	-12.25	Questionable reliability (R^2_L < 0.20).
hyp.L3	Selenium	109	27	29.59	0.00	0.24	8.36	5.65	
hyp.L3	Silver	215	46	48.88	0.00	0.22	3.90	1.29	
hyp.L3	Zinc	185	16	34.10	0.00	0.31	5.36	-14.72	
hyp.L3	Butyltin	50	4	10.96	0.00	0.39	3.30	-6.84	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L3	Dibutyltin	55	6	14.41	0.00	0.38	3.01	-7.54	
hyp.L3	Tributyltin	54	6	12.66	0.00	0.34	2.28	-7.13	
hyp.L3	Acenaphthene	161	10	40.22	0.00	0.54	1.83	-7.82	
hyp.L3	Anthracene	163	11	43.70	0.00	0.54	2.04	-8.43	
hyp.L3	Fluorene	158	9	39.03	0.00	0.57	2.10	-8.57	
hyp.L3	2-methylnaphthalene	158	7	34.45	0.00	0.60	1.90	-7.47	
hyp.L3	Acenaphthylene	167	13	43.17	0.00	0.47	2.22	-7.14	
hyp.L3	Naphthalene	132	12	45.32	0.00	0.56	2.40	-8.29	
hyp.L3	Phenanthrene	170	10	42.74	0.00	0.56	2.05	-10.02	
hyp.L3	Benzo(a)anthracene	173	13	44.39	0.00	0.48	2.15	-9.19	
hyp.L3	Benzo(a)pyrene	172	13	46.82	0.00	0.51	2.30	-10.02	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hyp.L3	Benzo(b)fluoranthene	171	13	44.15	0.00	0.48	2.28	-9.96	
hyp.L3	Benzo(ghi)perylene	173	13	48.50	0.00	0.53	2.43	-10.29	
hyp.L3	Benzo(k)fluoranthene	169	13	44.21	0.00	0.48	2.29	-8.87	
hyp.L3	Chrysene	170	13	43.89	0.00	0.48	2.21	-9.64	
hyp.L3	Dibenzanthracene	173	13	43.53	0.00	0.47	2.33	-7.97	
hyp.L3	Fluoranthene	175	14	46.67	0.00	0.48	2.09	-9.77	
hyp.L3	Indeno(c,d)pyrene	172	13	47.71	0.00	0.52	2.41	-10.15	
hyp.L3	Pyrene	174	13	46.87	0.00	0.51	2.10	-10.03	
hyp.L3	Total LPAH	171	11	44.72	0.00	0.55	2.03	-10.28	
hyp.L3	Total HPAH	176	13	46.59	0.00	0.50	2.23	-12.01	
hyp.L3	Total PAHs	177	13	47.05	0.00	0.51	2.15	-11.91	
hyp.L3	Diesel-range hydrocarbons	120	20	45.06	0.00	0.42	3.05	-9.35	
hyp.L3	Residual organics	109	17	38.81	0.00	0.41	4.08	-14.01	
hyp.L3	Dibenzofuran	161	11	35.92	0.00	0.45	1.93	-6.36	
hyp.L3	Hexachlorobenzene	94	5	12.28	0.00	0.31	2.60	-3.86	
hyp.L3	1,2,3,7,8-Pentachlorodibenzofuran	35	0	0.00	0.98	######	0.00	-11.20	Exclude (chi.p ≥ 0.01)
hyp.L3	Pentachlorodibenzo-p-dioxin homologs	41	1	8.04	0.00	0.86	86.28	-77.02	Exclude (only 1 hit retained)
hyp.L3	TEQ mammal (0.5 detection limit)	52	2	6.87	0.01	0.41	2.39	-4.87	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L3	Total dioxins/furans	52	2	3.93	0.05	0.23	1.80	-7.93	Exclude (chi.p ≥ 0.01)
hyp.L3	Total PCBs	139	8	17.62	0.00	0.29	2.10	-7.87	
hyp.L3	Aldrin	44	2	6.22	0.01	0.38	1.86	-4.47	Exclude (chi.p ≥ 0.01)
hyp.L3	alpha-Hexachlorocyclohexane	41	2	14.01	0.00	0.88	67.62	-31.53	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L3	beta-Hexachlorocyclohexane	80	13	22.41	0.00	0.32	3.38	-2.91	
hyp.L3	delta-Hexachlorocyclohexane	31	2	4.75	0.03	0.32	7.05	-0.61	Exclude (chi.p ≥ 0.01)
hyp.L3	Carbazole	122	8	35.01	0.00	0.59	2.47	-8.69	
hyp.L3	Methoxychlor	37	3	7.35	0.01	0.35	4.84	-6.00	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L3	cis-Nonachlor	50	6	12.32	0.00	0.34	5.45	-1.91	
hyp.L3	trans-Nonachlor	65	2	0.94	0.33	0.05	1.53	-2.99	Exclude (chi.p ≥ 0.01)
hyp.L3	Total chlordane	142	10	29.51	0.00	0.41	2.66	-4.38	
hyp.L3	DDD	178	15	53.89	0.00	0.52	2.63	-6.78	
hyp.L3	DDE	168	12	40.91	0.00	0.47	2.68	-5.69	
hyp.L3	DDT	147	5	19.23	0.00	0.44	1.77	-6.19	
hyp.L3	Total DDTs	177	12	45.48	0.00	0.52	2.47	-7.65	
hyp.L3	Total endosulfans	37	1	3.63	0.06	0.40	2.00	-4.89	Exclude (chi.p ≥ 0.01)
hyp.L3	4-Methylphenol	56	5	7.77	0.01	0.23	2.41	-6.43	
hyp.L3	Pentachlorophenol	42	5	8.17	0.00	0.27	2.55	-5.68	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hyp.L3	Phenol	53	9	14.60	0.00	0.30	3.68	-6.45	
hyp.L3	bis(2-ethylhexyl) phthalate	117	1	3.52	0.06	0.31	1.61	-9.56	Exclude (chi.p ≥ 0.01)
hyp.L3	Butylbenzyl phthalate	52	5	10.79	0.00	0.33	2.68	-7.88	
hyp.L3	Dibutyl phthalate	72	9	18.19	0.00	0.34	2.64	-6.31	
hyp.L2	Ammonia	203	86	73.34	0.00	0.27	5.30	-10.47	
hyp.L2	Sulfide	133	36	63.64	0.00	0.41	3.19	-4.46	
hyp.L2	Fines (%)	218	101	110.69	0.00	0.37	7.06	-11.98	
hyp.L2	Aluminum	209	92	79.64	0.00	0.28	12.72	-55.57	
hyp.L2	Antimony	102	19	43.56	0.00	0.44	4.24	-1.04	
hyp.L2	Arsenic	169	52	43.03	0.00	0.21	5.81	-4.71	
hyp.L2	Cadmium	175	59	60.38	0.00	0.27	5.01	1.78	
hyp.L2	Chromium	183	66	36.73	0.00	0.15	5.44	-8.88	Questionable reliability ($R^2_L < 0.20$).
hyp.L2	Copper	156	39	56.12	0.00	0.32	3.88	-8.06	
hyp.L2	Lead	149	32	52.50	0.00	0.34	4.12	-7.70	
hyp.L2	Mercury	162	49	60.25	0.00	0.30	4.72	3.82	
hyp.L2	Nickel	180	72	41.53	0.00	0.17	9.81	-14.20	Questionable reliability ($R^2_L < 0.20$).
hyp.L2	Selenium	99	49	54.19	0.00	0.39	11.65	9.75	
hyp.L2	Silver	197	80	76.76	0.00	0.29	5.05	3.18	
hyp.L2	Zinc	158	41	56.70	0.00	0.31	5.85	-14.17	
hyp.L2	Butyltin	36	8	9.78	0.00	0.26	2.19	-3.91	
hyp.L2	Dibutyltin	42	12	18.69	0.00	0.37	3.06	-6.24	
hyp.L2	Tributyltin	41	12	23.89	0.00	0.48	4.06	-9.95	
hyp.L2	Acenaphthene	114	11	39.02	0.00	0.54	1.83	-7.51	
hyp.L2	Anthracene	116	13	41.95	0.00	0.52	1.93	-7.60	
hyp.L2	Fluorene	112	11	38.54	0.00	0.54	1.99	-7.72	
hyp.L2	2-methylnaphthalene	108	6	32.21	0.00	0.69	2.17	-8.59	
hyp.L2	Acenaphthylene	118	14	41.55	0.00	0.48	2.20	-6.76	
hyp.L2	Naphthalene	95	14	43.91	0.00	0.55	2.44	-7.87	
hyp.L2	Phenanthrene	120	11	42.33	0.00	0.58	2.11	-9.96	
hyp.L2	Benzo(a)anthracene	121	13	42.57	0.00	0.52	2.26	-9.44	
hyp.L2	Benzo(a)pyrene	122	15	46.15	0.00	0.51	2.29	-9.57	
hyp.L2	Benzo(b)fluoranthene	120	14	42.88	0.00	0.50	2.31	-9.78	
hyp.L2	Benzo(ghi)perylene	123	15	48.30	0.00	0.53	2.43	-9.87	
hyp.L2	Benzo(k)fluoranthene	119	14	42.88	0.00	0.50	2.32	-8.68	
hyp.L2	Chrysene	118	13	42.05	0.00	0.51	2.32	-9.90	
hyp.L2	Dibenzanthracene	124	15	43.26	0.00	0.47	2.31	-7.51	

Table E-1. Results for Individual LRMs

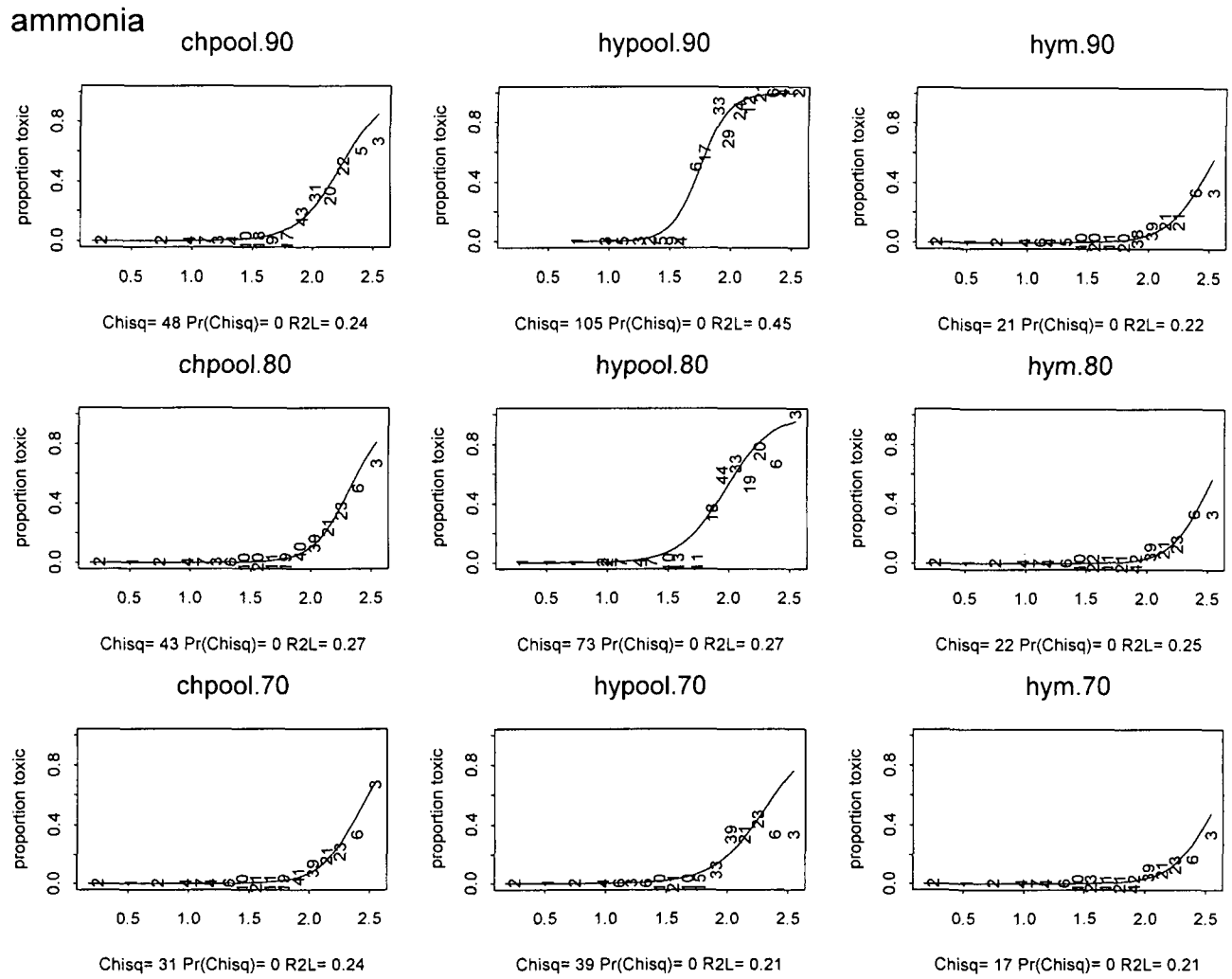
Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hyp.L2	Fluoranthene	121	12	43.71	0.00	0.56	2.39	-11.16	
hyp.L2	Indeno(c,d)pyrene	122	15	47.28	0.00	0.52	2.40	-9.71	
hyp.L2	Pyrene	121	12	44.47	0.00	0.57	2.32	-11.00	
hyp.L2	Total LPAH	121	12	43.28	0.00	0.55	2.05	-10.07	
hyp.L2	Total HPAH	125	14	45.54	0.00	0.52	2.27	-11.90	
hyp.L2	Total PAHs	125	13	45.14	0.00	0.54	2.26	-12.32	
hyp.L2	Diesel-range hydrocarbons	88	25	43.30	0.00	0.41	3.13	-8.79	
hyp.L2	Residual organics	79	24	39.24	0.00	0.40	4.38	-13.92	
hyp.L2	Dibenzofuran	114	13	34.46	0.00	0.43	1.84	-5.69	
hyp.L2	Hexachlorobenzene	67	4	12.14	0.00	0.40	2.71	-3.88	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L2	1,2,3,7,8-Pentachlorodibenzofuran	28	1	1.72	0.19	0.20	1.30	-3.73	Exclude (chi.p $\geq$ 0.01)
hyp.L2	Pentachlorodibenzo-p-dioxin homologs	37	5	13.97	0.00	0.48	6.38	-4.81	
hyp.L2	TEQ mammal (0.5 detection limit)	44	6	13.58	0.00	0.39	2.31	-2.93	
hyp.L2	Total dioxins/furans	48	10	20.31	0.00	0.41	3.28	-9.81	
hyp.L2	Total PCBs	114	25	48.39	0.00	0.40	2.67	-7.45	
hyp.L2	Aldrin	36	4	11.30	0.00	0.45	2.17	-3.49	Questionable reliability (fewer than 5 toxic stations retained)
hyp.L2	alpha-Hexachlorocyclohexane	33	9	16.57	0.00	0.43	4.39	-0.07	
hyp.L2	beta-Hexachlorocyclohexane	73	28	44.00	0.00	0.45	4.93	-1.77	
hyp.L2	delta-Hexachlorocyclohexane	26	7	7.26	0.01	0.24	5.34	1.08	
hyp.L2	Carbazole	93	10	37.84	0.00	0.60	2.64	-8.48	
hyp.L2	Methoxychlor	34	9	16.52	0.00	0.42	4.55	-3.72	
hyp.L2	cis-Nonachlor	45	14	20.78	0.00	0.37	5.63	-0.30	
hyp.L2	trans-Nonachlor	53	9	9.78	0.00	0.20	2.95	-0.98	
hyp.L2	Total chlordane	103	12	41.60	0.00	0.56	3.73	-4.67	
hyp.L2	DDD	134	21	60.10	0.00	0.52	2.60	-5.85	
hyp.L2	DDE	127	18	50.25	0.00	0.48	2.85	-4.96	
hyp.L2	DDT	104	6	26.25	0.00	0.57	2.23	-6.85	
hyp.L2	Total DDTs	131	16	53.73	0.00	0.55	2.65	-7.36	
hyp.L2	Total endosulfans	31	2	6.11	0.01	0.41	1.96	-3.74	Exclude (chi.p $\geq$ 0.01)
hyp.L2	4-Methylphenol	44	14	18.36	0.00	0.33	3.12	-5.85	
hyp.L2	Pentachlorophenol	33	7	10.04	0.00	0.29	2.38	-4.61	
hyp.L2	Phenol	47	18	19.76	0.00	0.32	4.84	-6.45	
hyp.L2	bis(2-ethylhexyl) phthalate	78	2	6.18	0.01	0.33	1.59	-3.34	Exclude (chi.p $\geq$ 0.01)
hyp.L2	Butylbenzyl phthalate	40	10	14.75	0.00	0.33	2.71	-6.66	
hyp.L2	Dibutyl phthalate	64	17	30.35	0.00	0.41	3.42	-6.42	
hyp.L1	Ammonia	187	128	104.78	0.00	0.45	8.53	-14.95	

Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hyp.L1	Sulfide	118	73	84.00	0.00	0.54	5.36	-4.26	
hyp.L1	Fines (%)	196	137	131.26	0.00	0.55	8.43	-12.49	
hyp.L1	Aluminum	193	134	79.31	0.00	0.33	13.66	-58.12	
hyp.L1	Antimony	79	37	46.72	0.00	0.43	5.38	1.07	
hyp.L1	Arsenic	149	90	53.55	0.00	0.27	9.27	-5.39	
hyp.L1	Cadmium	147	88	56.69	0.00	0.29	5.75	3.51	
hyp.L1	Chromium	134	75	38.61	0.00	0.21	6.69	-9.97	
hyp.L1	Copper	159	100	75.22	0.00	0.36	6.54	-10.16	
hyp.L1	Lead	129	70	63.68	0.00	0.36	5.60	-8.04	
hyp.L1	Mercury	146	91	71.17	0.00	0.37	6.15	7.08	
hyp.L1	Nickel	163	112	47.45	0.00	0.23	14.50	-19.34	
hyp.L1	Selenium	85	67	35.34	0.00	0.40	11.91	11.94	
hyp.L1	Silver	168	109	89.45	0.00	0.41	7.36	6.06	
hyp.L1	Zinc	126	67	54.74	0.00	0.31	6.68	-14.54	
hyp.L1	Butyltin	26	13	12.44	0.00	0.35	2.73	-3.00	
hyp.L1	Dibutyltin	32	19	18.78	0.00	0.43	3.67	-5.53	
hyp.L1	Tributyltin	30	17	21.73	0.00	0.53	4.98	-10.41	
hyp.L1	Acenaphthene	72	21	55.22	0.00	0.64	2.44	-7.35	
hyp.L1	Anthracene	72	21	53.44	0.00	0.61	2.58	-7.97	
hyp.L1	Fluorene	73	23	55.68	0.00	0.61	2.58	-7.13	
hyp.L1	2-methylnaphthalene	79	28	60.22	0.00	0.59	3.34	-6.73	
hyp.L1	Acenaphthylene	71	20	45.98	0.00	0.54	2.61	-6.44	
hyp.L1	Naphthalene	67	29	49.41	0.00	0.54	3.47	-7.95	
hyp.L1	Phenanthrene	77	23	57.52	0.00	0.61	2.60	-9.48	
hyp.L1	Benzo(a)anthracene	75	21	52.20	0.00	0.59	2.68	-9.27	
hyp.L1	Benzo(a)pyrene	71	18	52.66	0.00	0.65	3.15	-11.75	
hyp.L1	Benzo(b)fluoranthene	74	21	51.41	0.00	0.58	2.82	-10.09	
hyp.L1	Benzo(ghi)perylene	72	19	53.92	0.00	0.65	3.17	-11.38	
hyp.L1	Benzo(k)fluoranthene	75	22	51.57	0.00	0.57	2.77	-8.45	
hyp.L1	Chrysene	75	22	51.98	0.00	0.57	2.73	-9.56	
hyp.L1	Dibenzanthracene	78	24	53.65	0.00	0.56	2.92	-7.50	
hyp.L1	Fluoranthene	77	23	54.80	0.00	0.58	2.65	-10.08	
hyp.L1	Indeno(c,d)pyrene	72	19	53.25	0.00	0.64	3.11	-11.07	
hyp.L1	Pyrene	74	20	52.89	0.00	0.61	2.62	-10.39	
hyp.L1	Total LPAH	76	22	58.88	0.00	0.64	2.79	-11.07	
hyp.L1	Total HPAH	76	21	53.41	0.00	0.60	2.72	-12.44	

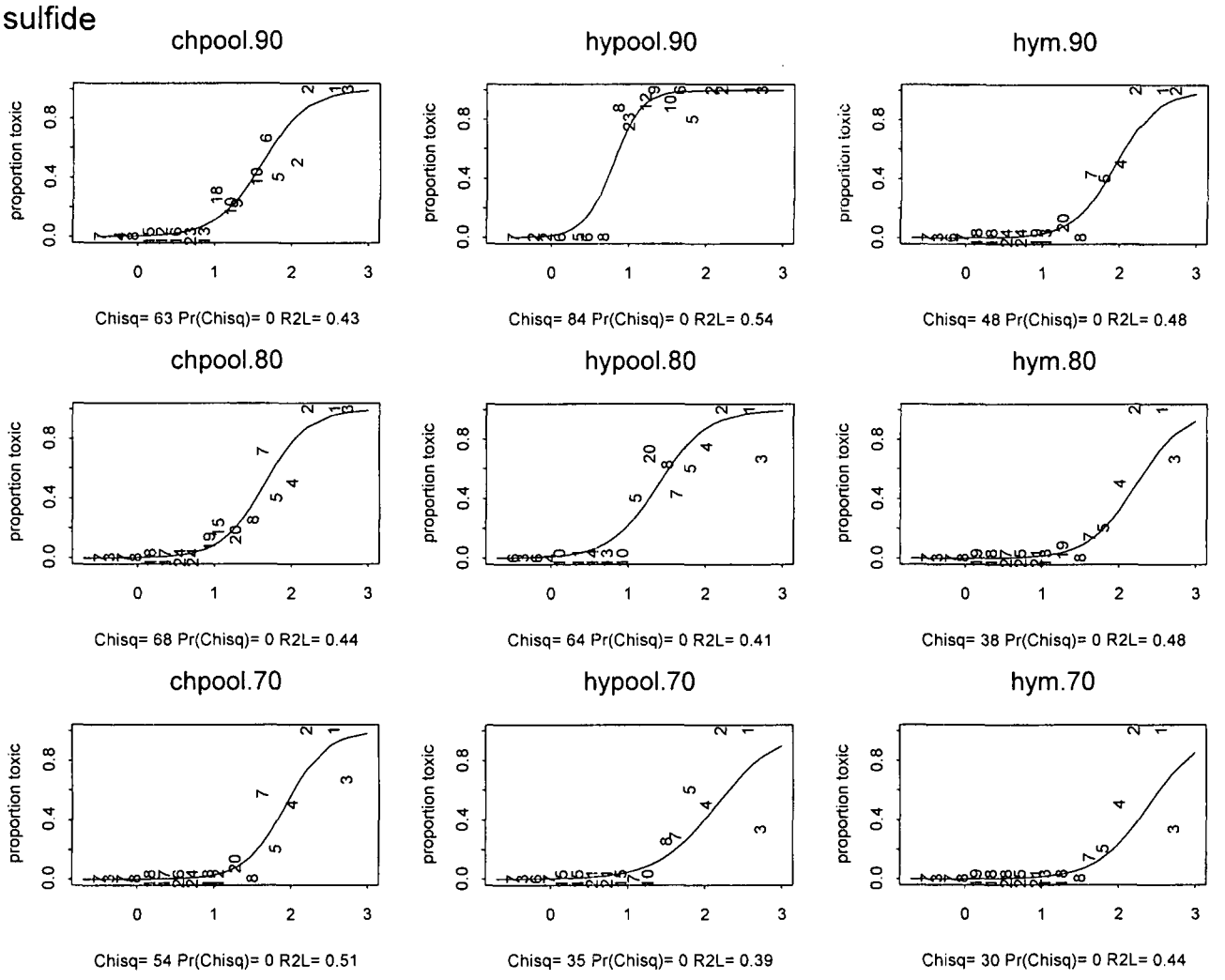
Table E-1. Results for Individual LRMs

Effect Level	Chemical	Screened Data Set							Comment
		# Samps Retained	# Toxic Retained	Chi-sq Statistic	Chi-sq p-value	R^2_L	LRM Slope	LRM Intercept	
hyp.L1	Total PAHs	79	23	56.82	0.00	0.60	2.74	-12.64	
hyp.L1	Diesel-range hydrocarbons	76	43	46.99	0.00	0.45	4.02	-9.20	
hyp.L1	Residual organics	78	51	36.61	0.00	0.36	4.82	-12.95	
hyp.L1	Dibenzofuran	80	29	56.55	0.00	0.54	2.73	-5.29	
hyp.L1	Hexachlorobenzene	67	33	36.92	0.00	0.40	4.24	-0.88	
hyp.L1	1,2,3,7,8-Pentachlorodibenzofuran	19	2	2.94	0.09	0.23	1.34	-2.70	Exclude (chi.p ≥ 0.01)
hyp.L1	Pentachlorodibenzo-p-dioxin homologs	27	8	20.79	0.00	0.63	7.80	-4.04	
hyp.L1	TEQ mammal (0.5 detection limit)	28	6	12.40	0.00	0.43	2.32	-2.47	
hyp.L1	Total dioxins/furans	34	12	19.14	0.00	0.43	3.35	-9.17	
hyp.L1	Total PCBs	74	29	44.05	0.00	0.44	3.01	-7.58	
hyp.L1	Aldrin	29	11	17.67	0.00	0.46	3.36	-1.48	
hyp.L1	alpha-Hexachlorocyclohexane	23	10	17.06	0.00	0.54	7.36	1.21	
hyp.L1	beta-Hexachlorocyclohexane	61	35	40.54	0.00	0.49	5.84	-1.00	
hyp.L1	delta-Hexachlorocyclohexane	24	14	9.68	0.00	0.30	7.30	3.66	
hyp.L1	Carbazole	69	27	44.27	0.00	0.48	2.67	-5.71	
hyp.L1	Methoxychlor	26	14	17.88	0.00	0.50	4.33	-1.56	
hyp.L1	cis-Nonachlor	35	19	20.04	0.00	0.42	5.94	0.98	
hyp.L1	trans-Nonachlor	41	19	17.49	0.00	0.31	4.53	1.16	
hyp.L1	Total chlordane	89	47	54.50	0.00	0.44	4.79	-1.93	
hyp.L1	DDD	90	33	67.49	0.00	0.57	3.20	-5.08	
hyp.L1	DDE	78	24	50.72	0.00	0.53	3.12	-3.97	
hyp.L1	DDT	70	23	45.64	0.00	0.51	2.75	-4.36	
hyp.L1	Total DDTs	93	35	65.76	0.00	0.53	3.10	-5.78	
hyp.L1	Total endosulfans	15	2	4.70	0.03	0.40	1.62	-2.70	Exclude (chi.p ≥ 0.01)
hyp.L1	4-Methylphenol	28	13	15.31	0.00	0.40	3.39	-5.89	
hyp.L1	Pentachlorophenol	42	30	25.08	0.00	0.50	5.87	-5.18	
hyp.L1	Phenol	27	16	14.89	0.00	0.41	5.57	-6.73	
hyp.L1	bis(2-ethylhexyl) phthalate	52	16	34.87	0.00	0.54	3.34	-10.87	
hyp.L1	Butylbenzyl phthalate	27	13	13.11	0.00	0.35	3.62	-7.76	
hyp.L1	Dibutyl phthalate	48	20	31.31	0.00	0.48	4.22	-7.03	



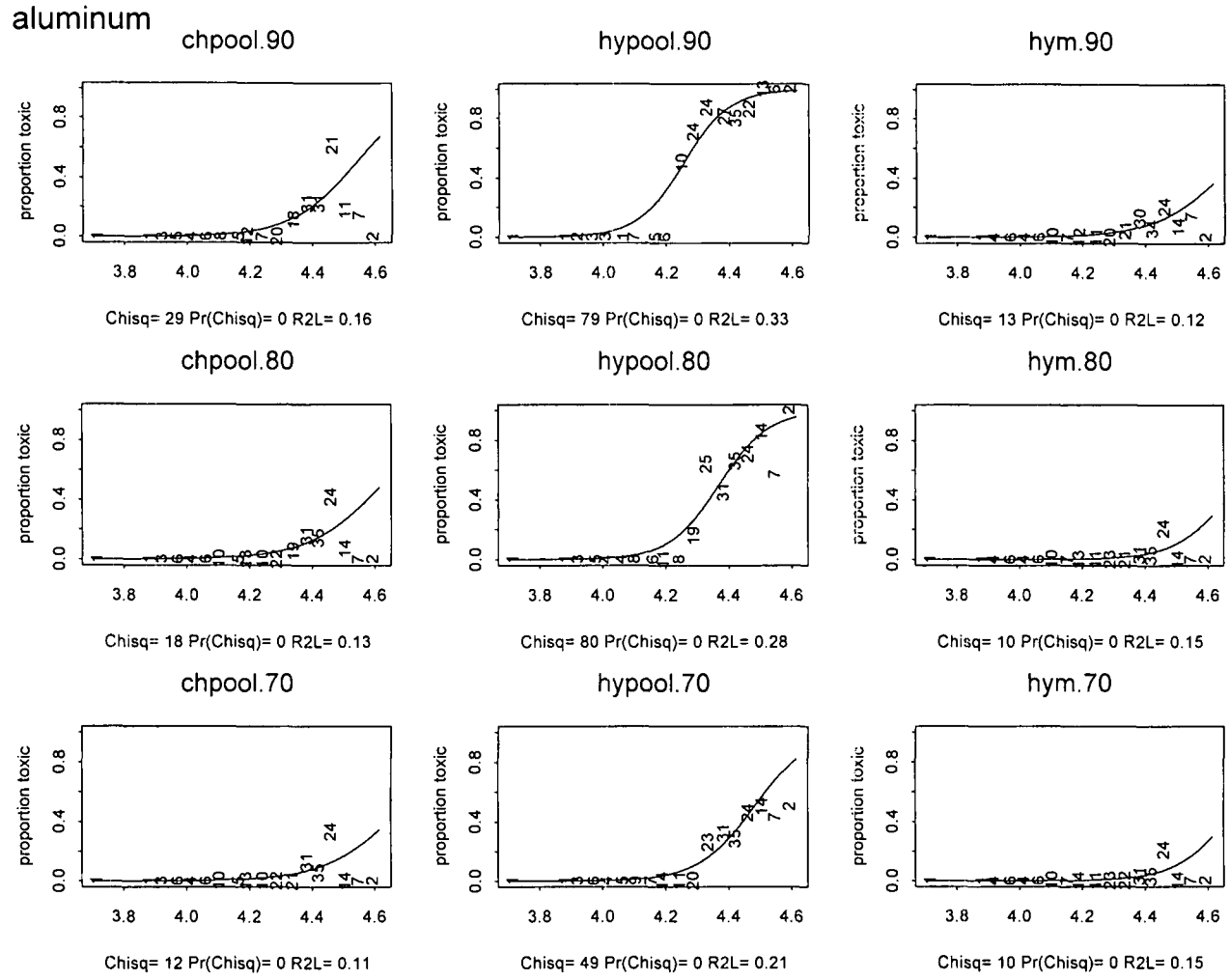
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-1. Logistic regression model – ammonia



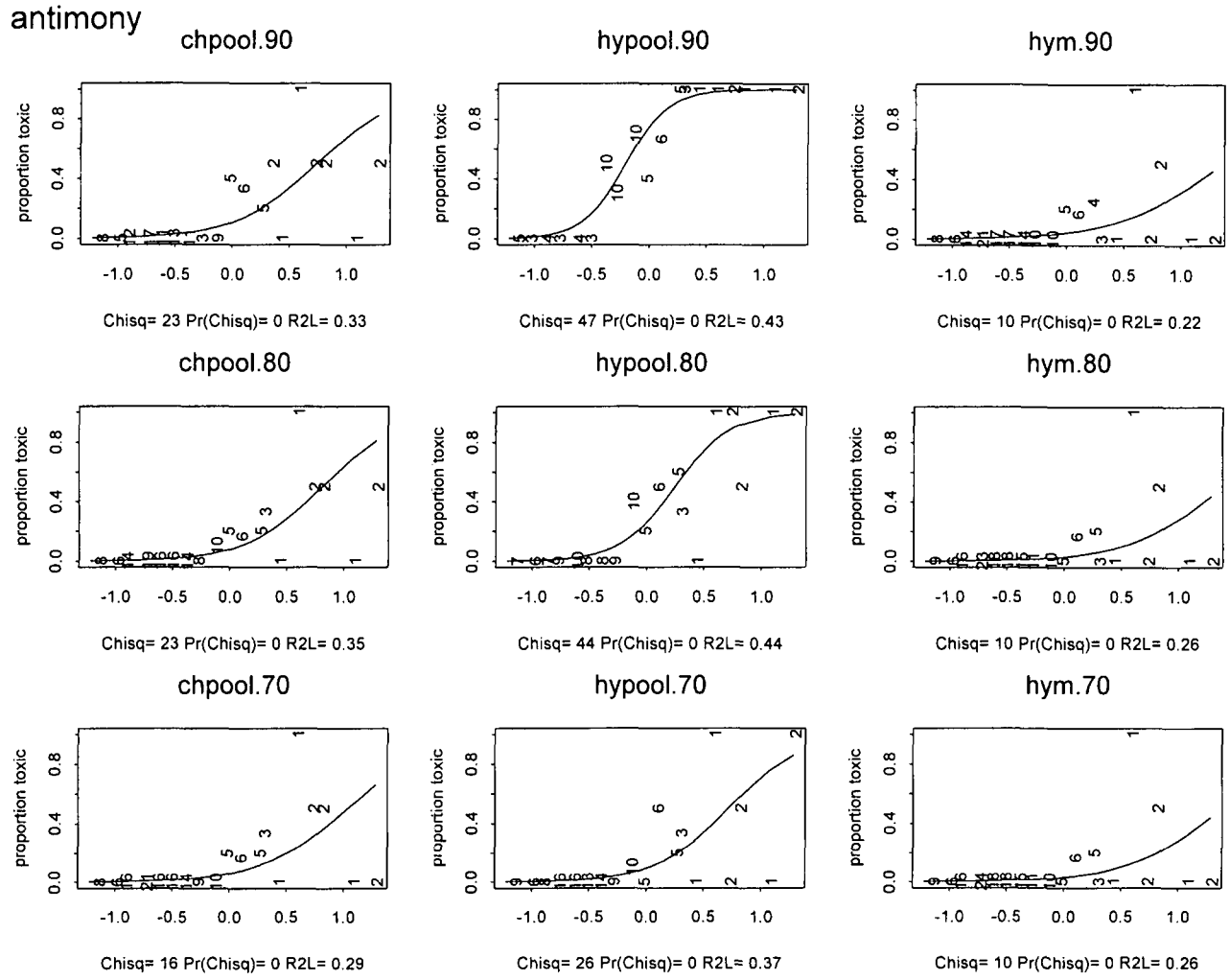
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-2. Logistic regression model – sulfide



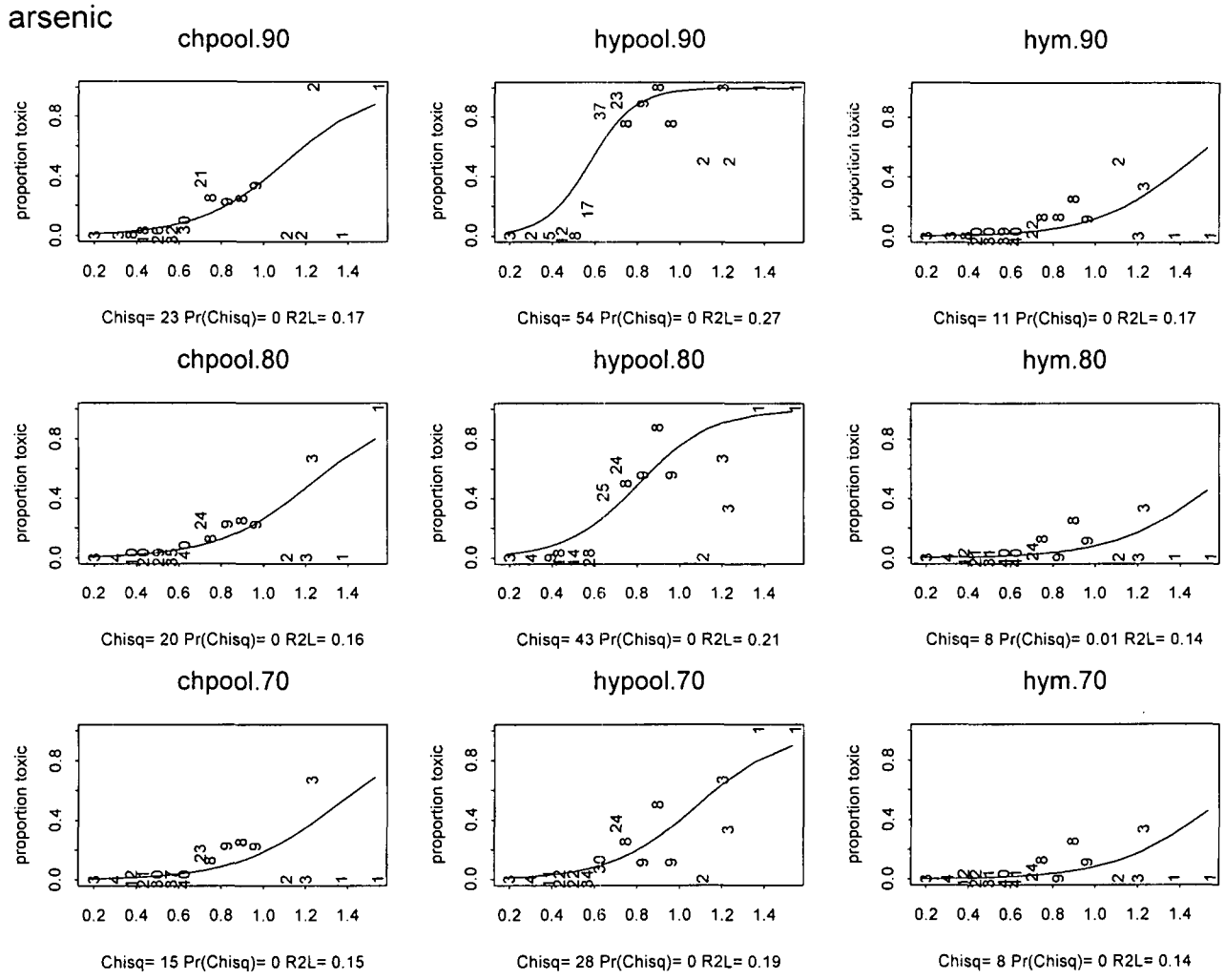
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-4. Logistic regression model – aluminum



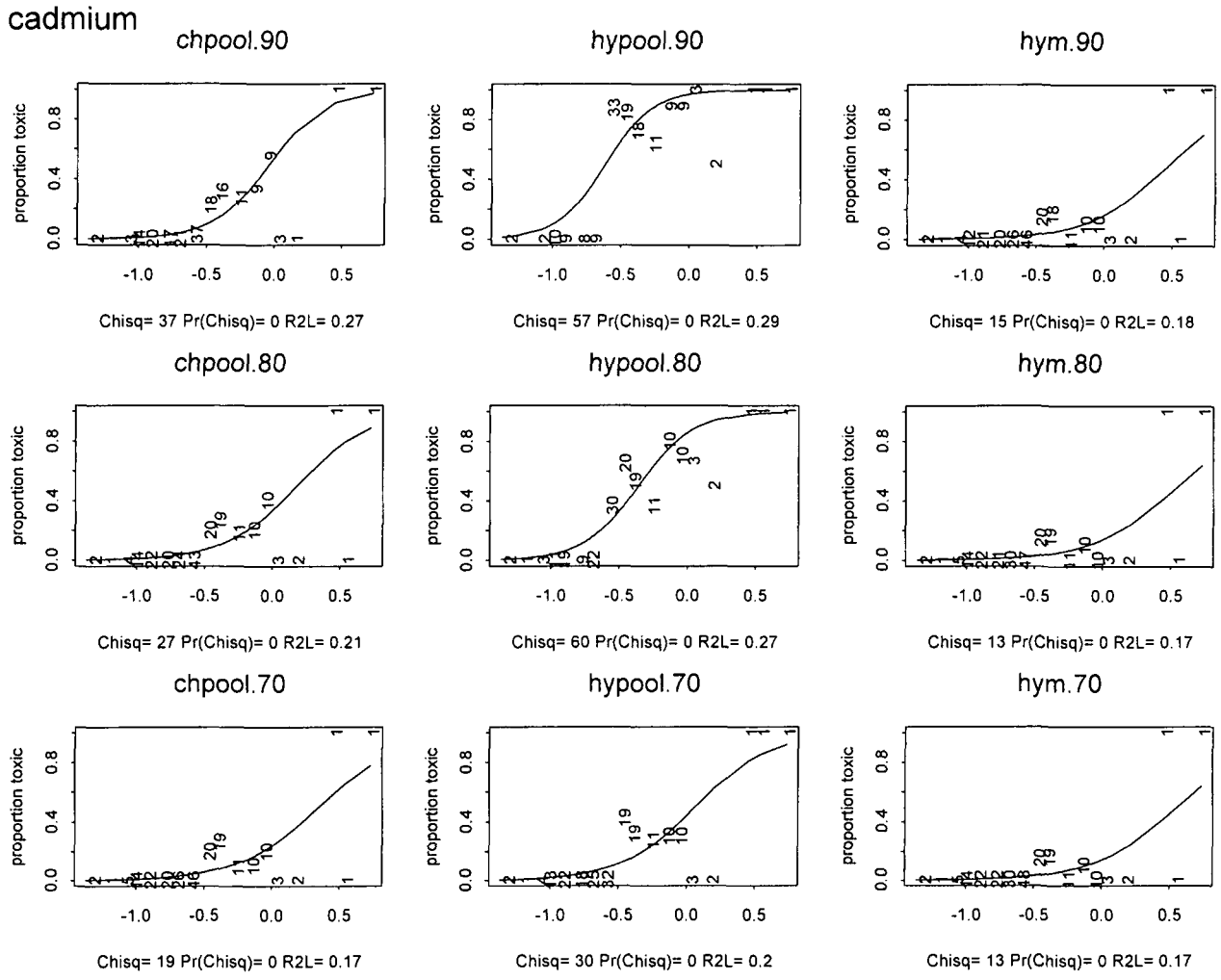
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-5. Logistic regression model – antimony



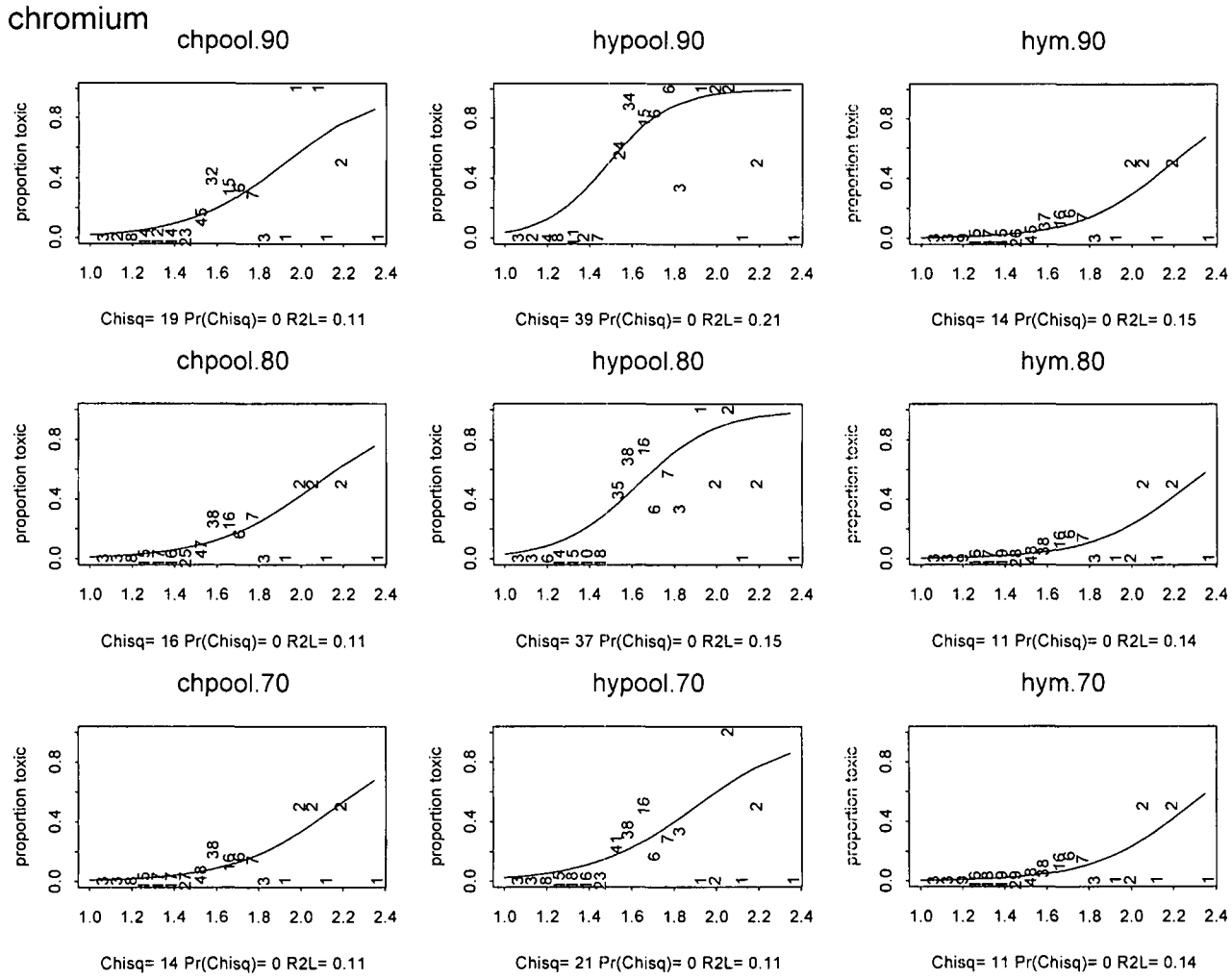
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-6. Logistic regression model – arsenic



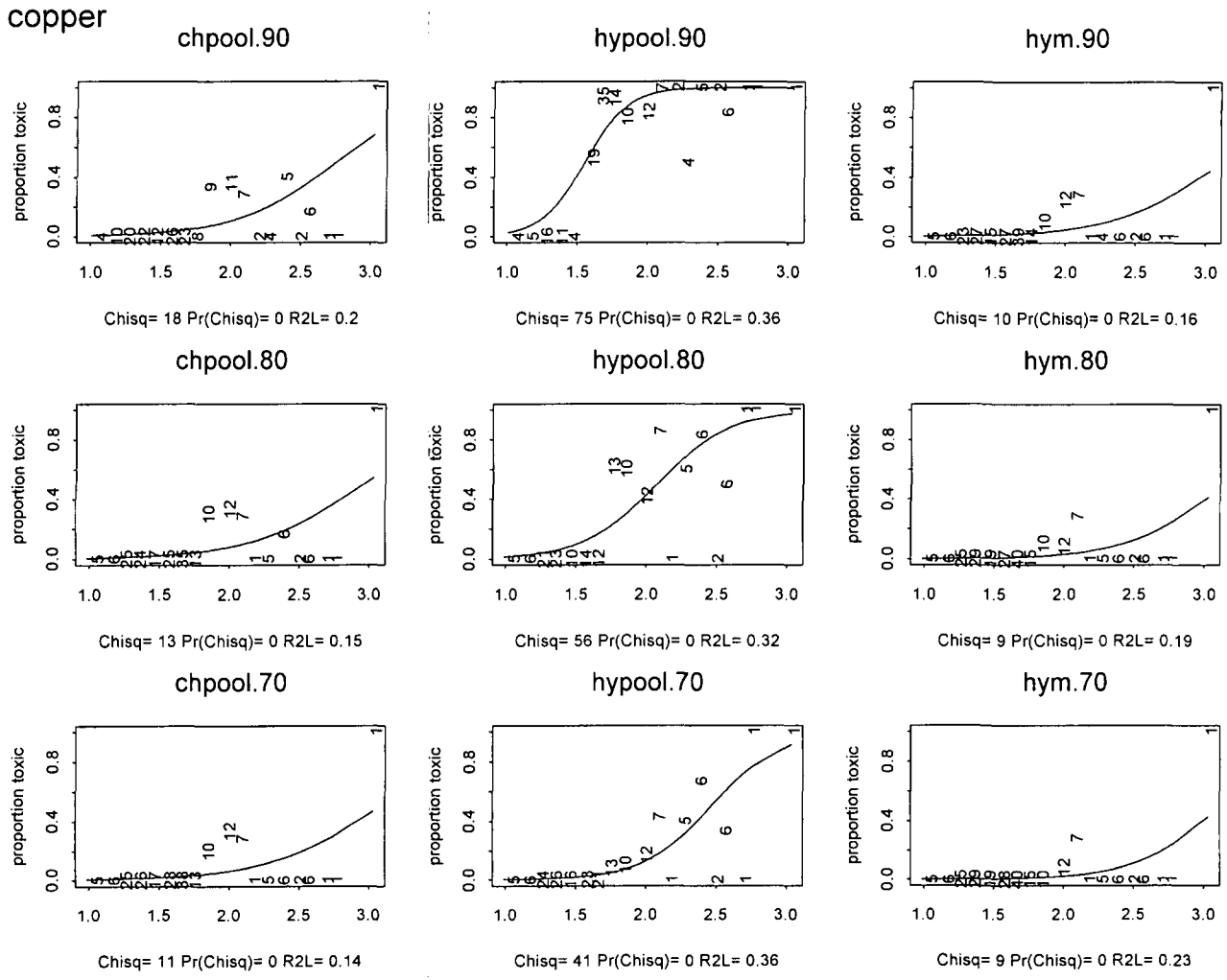
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-7. Logistic regression model – cadmium



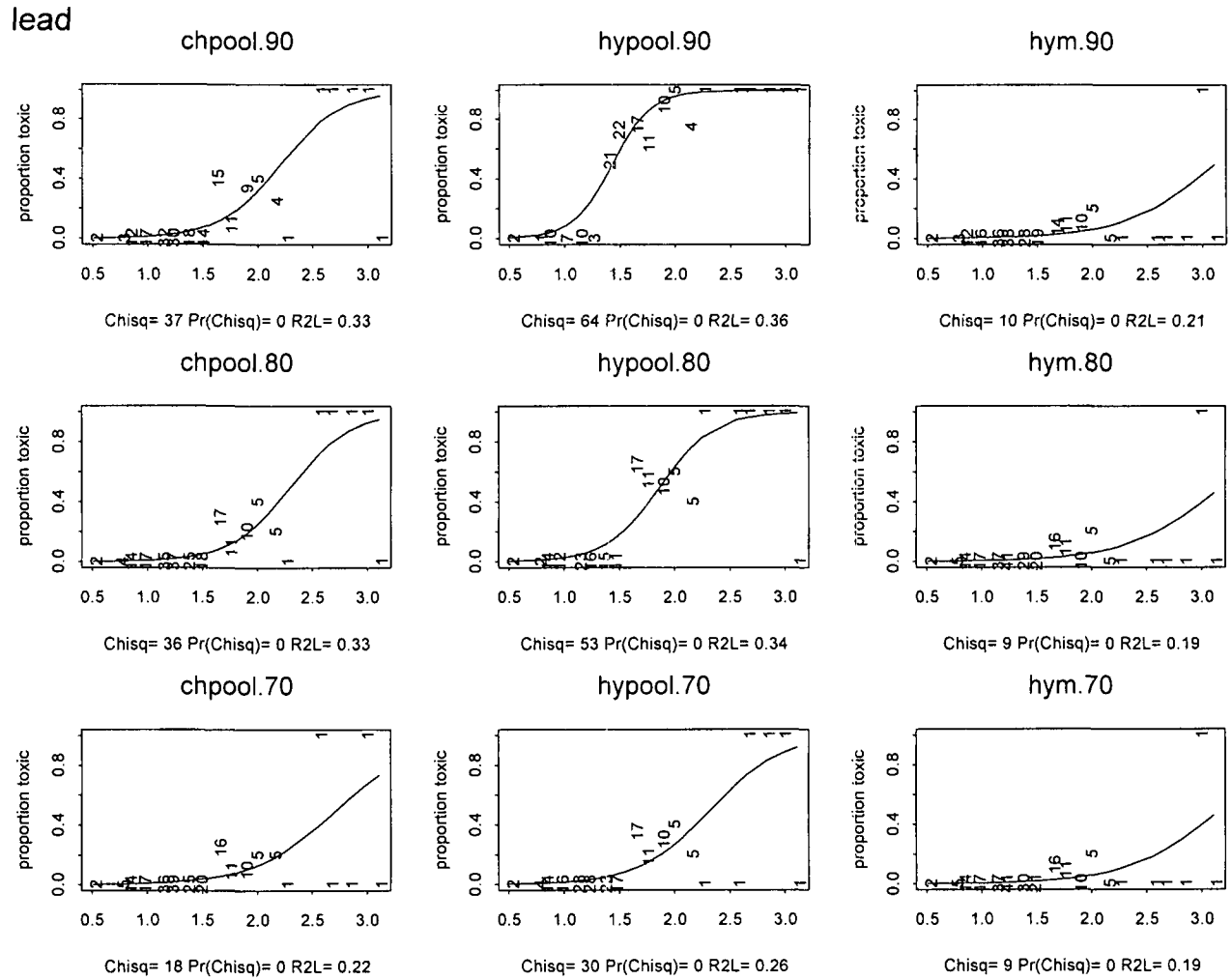
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-8. Logistic regression model – chromium



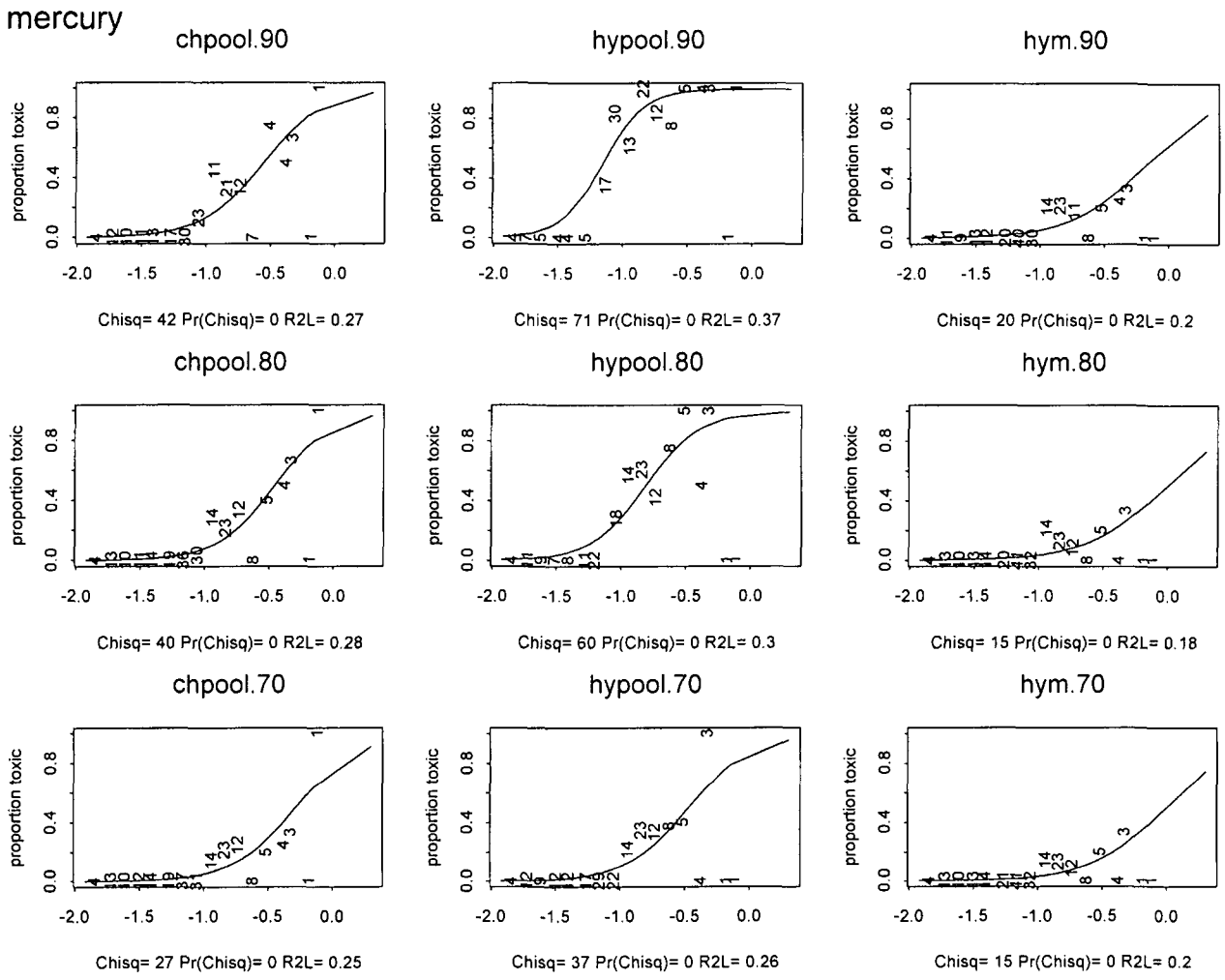
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-9. Logistic regression model – copper



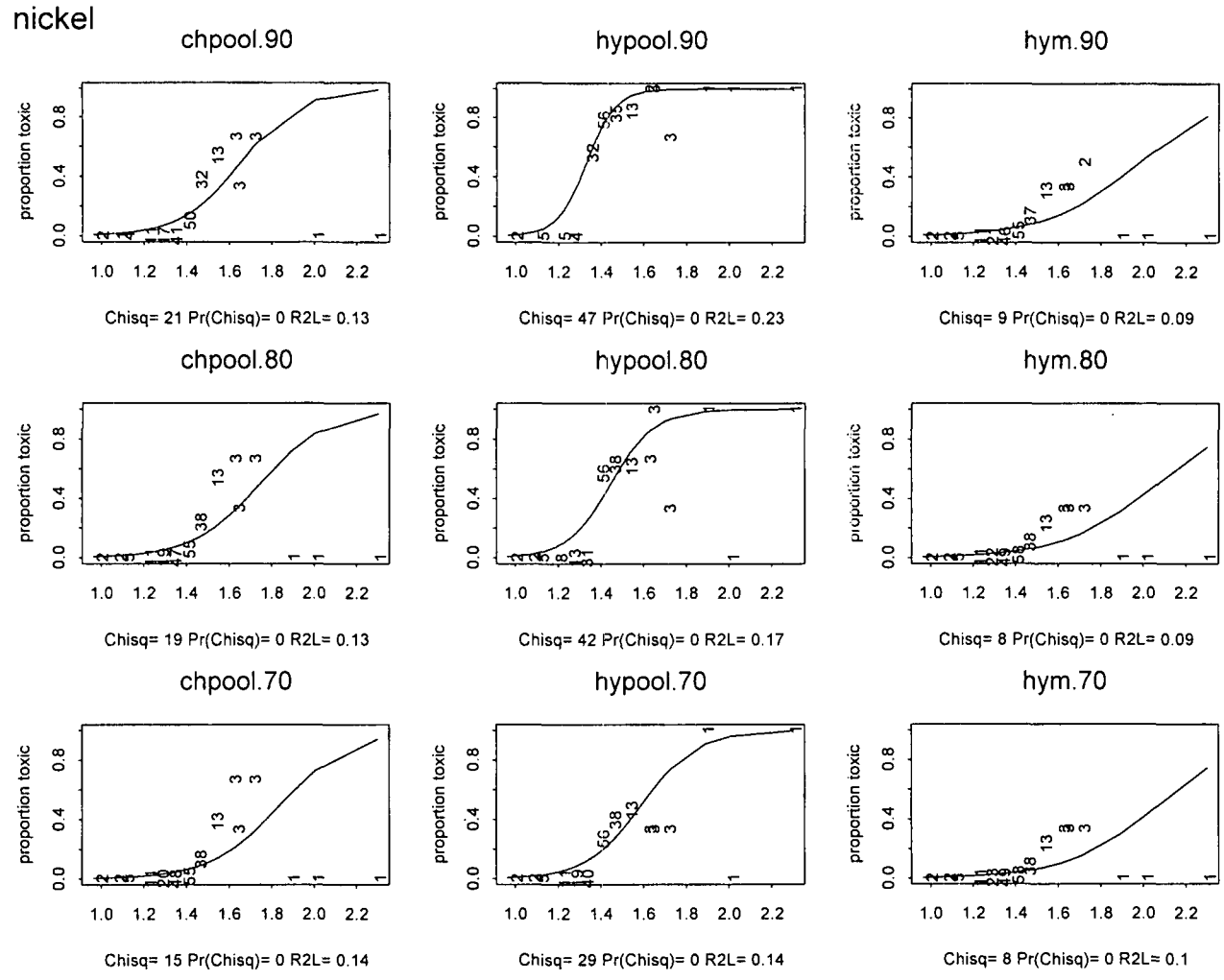
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-10. Logistic regression model – lead



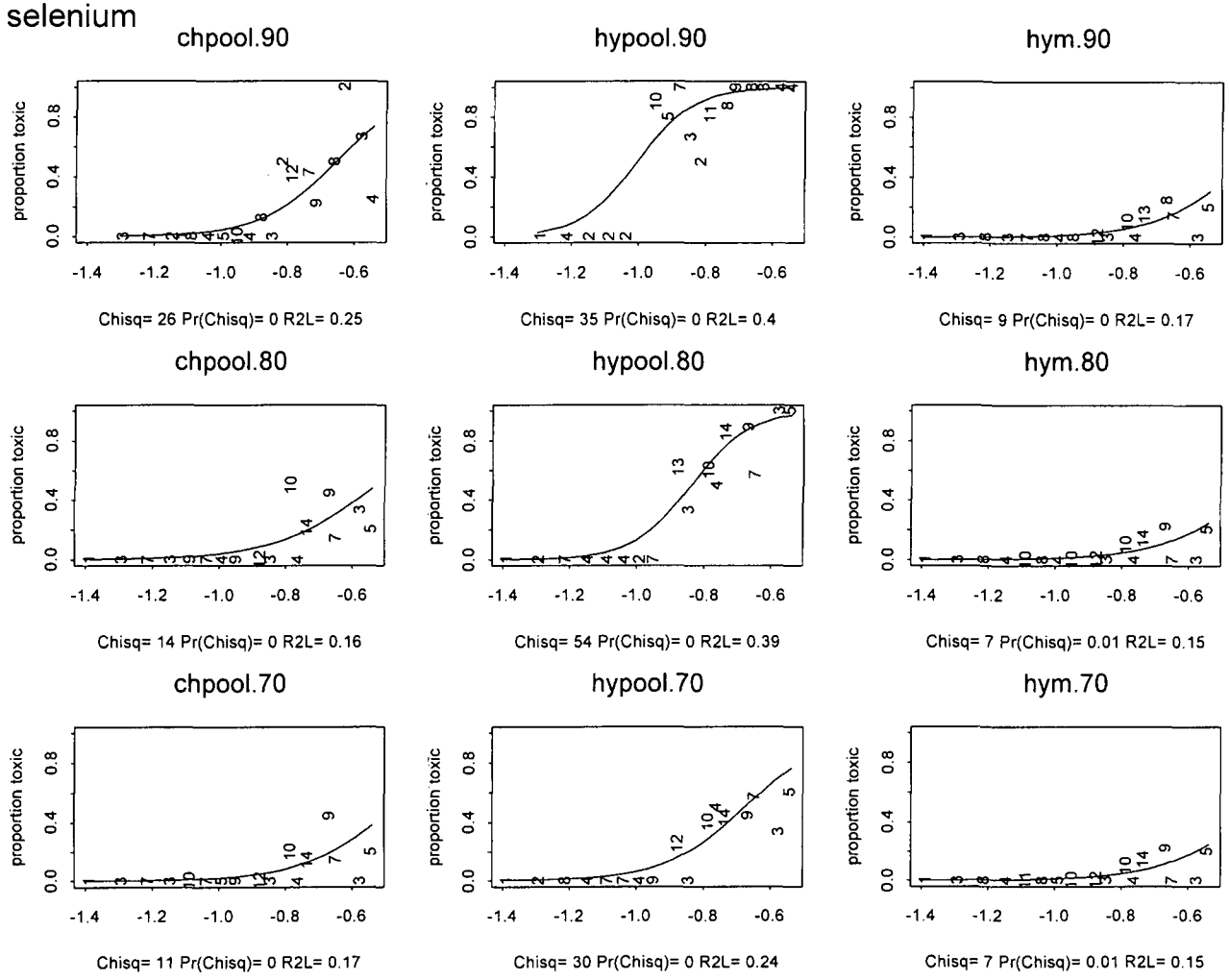
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-11. Logistic regression model – mercury



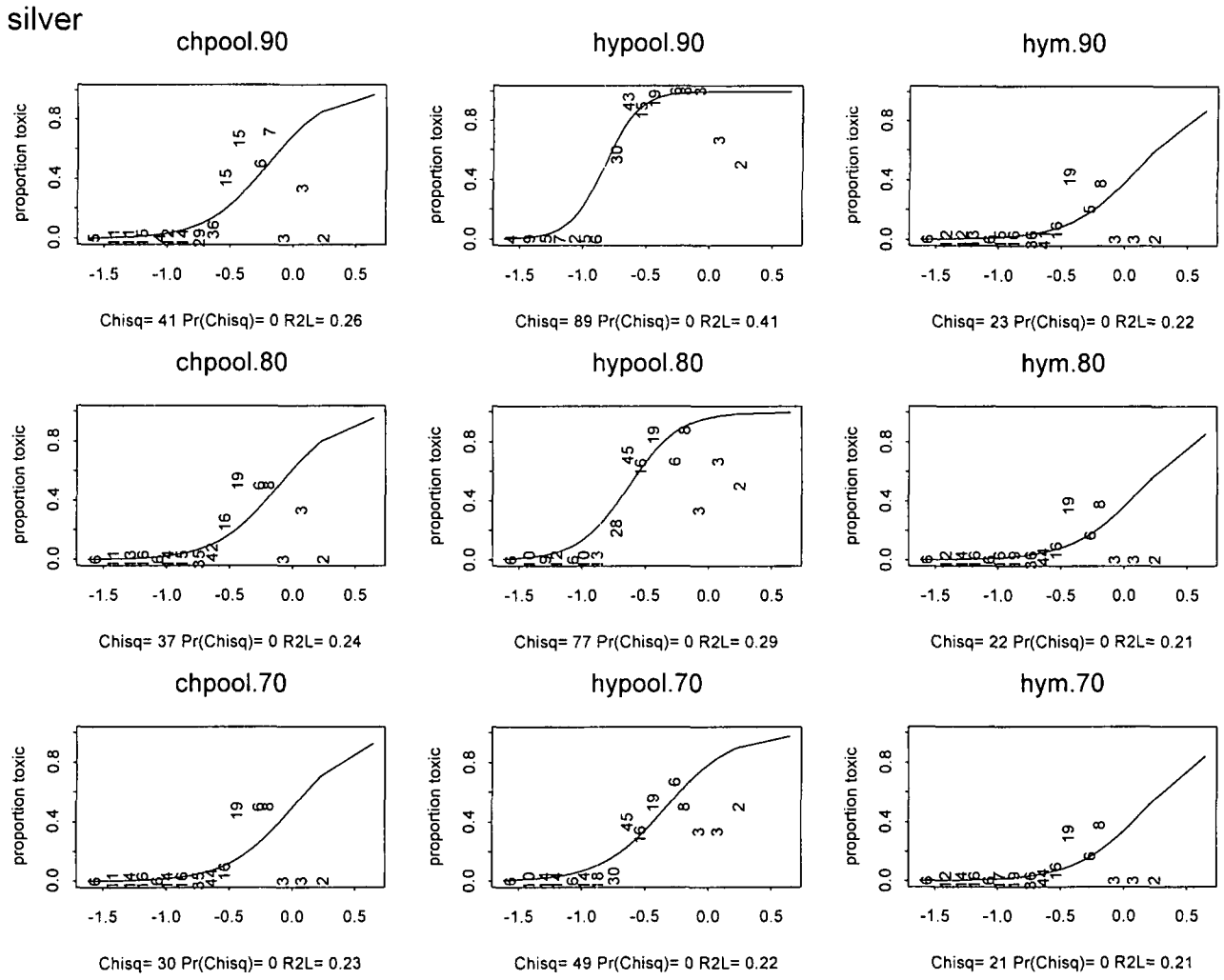
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-12. Logistic regression model – nickel



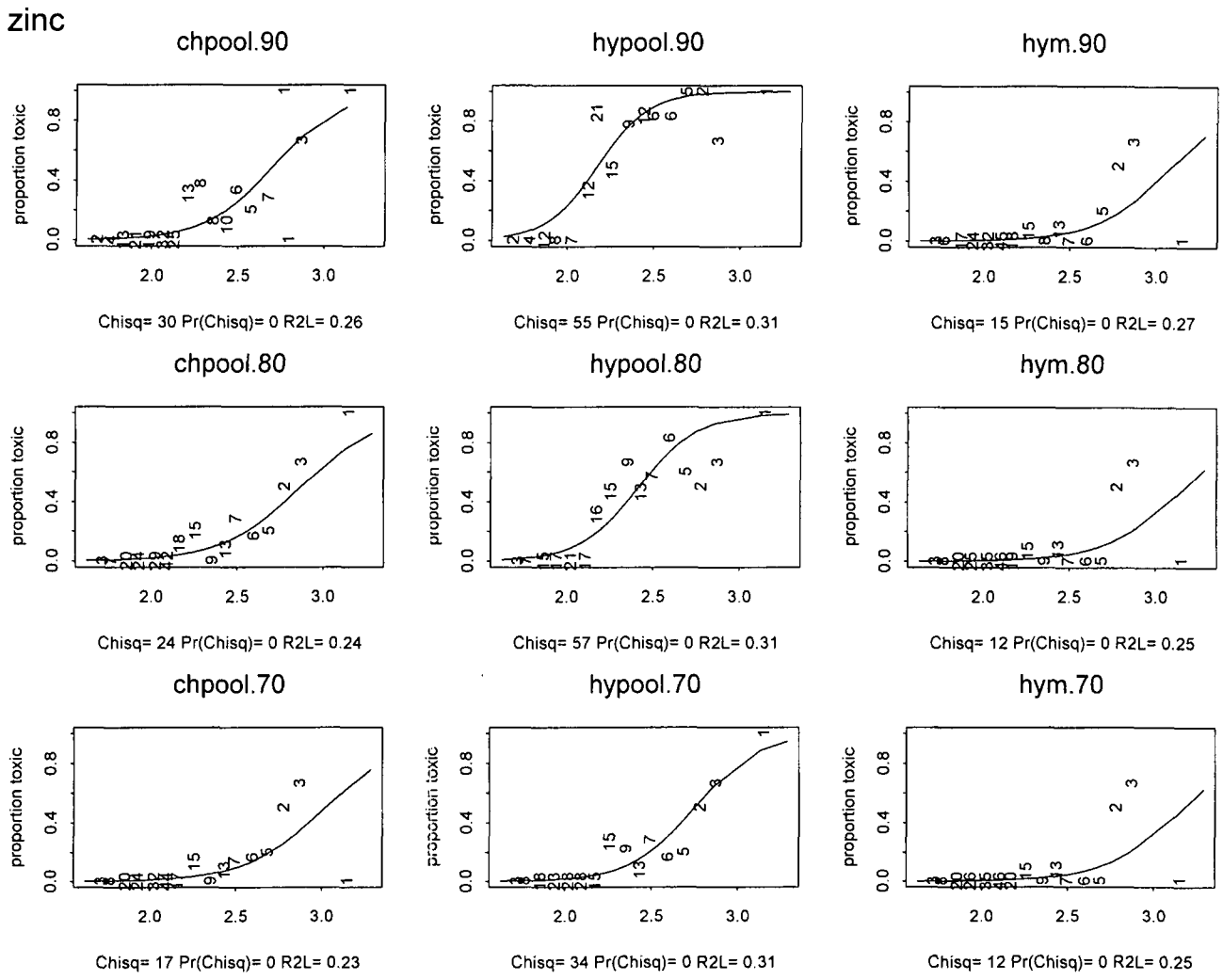
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-13. Logistic regression model – selenium



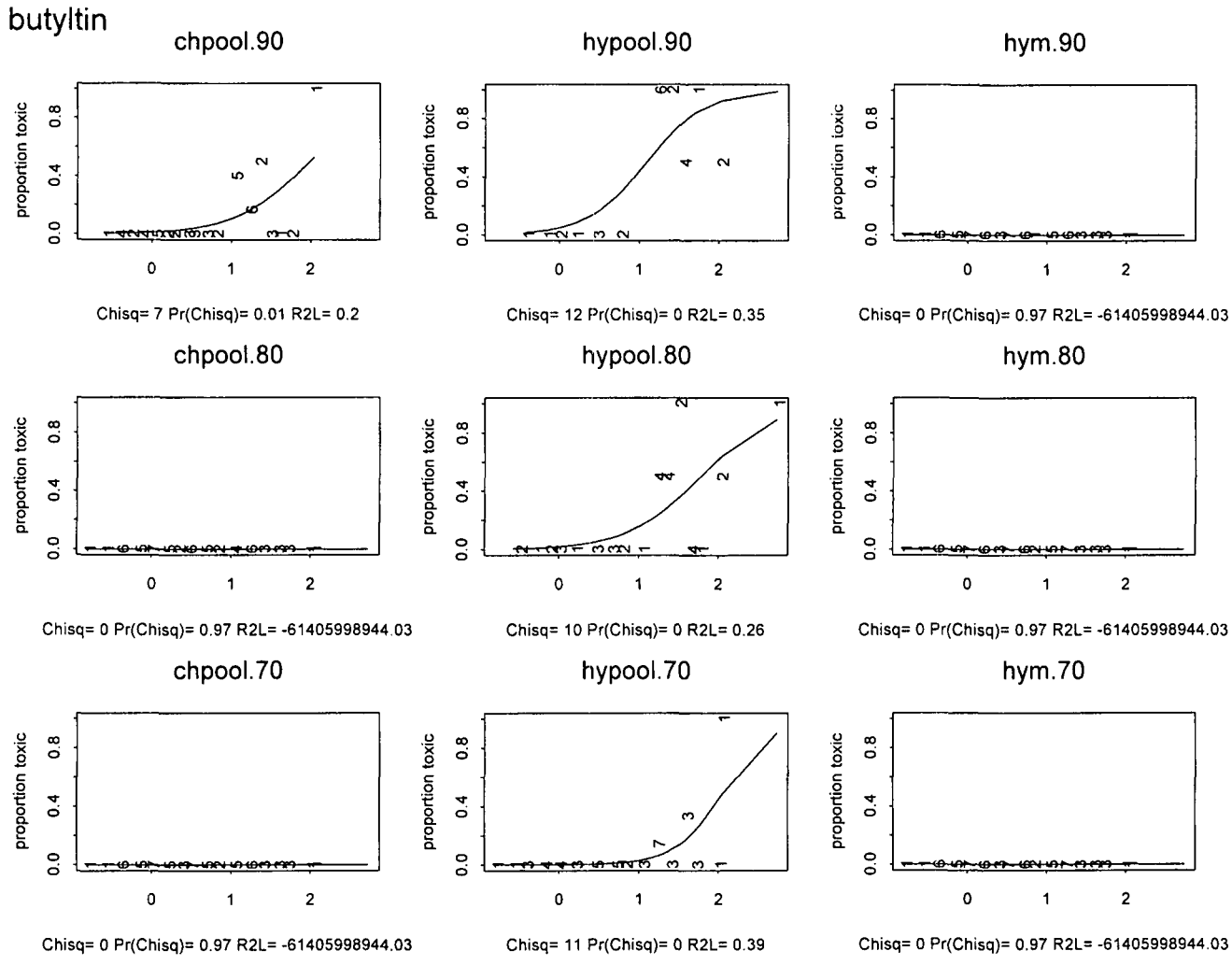
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-14. Logistic regression model – silver



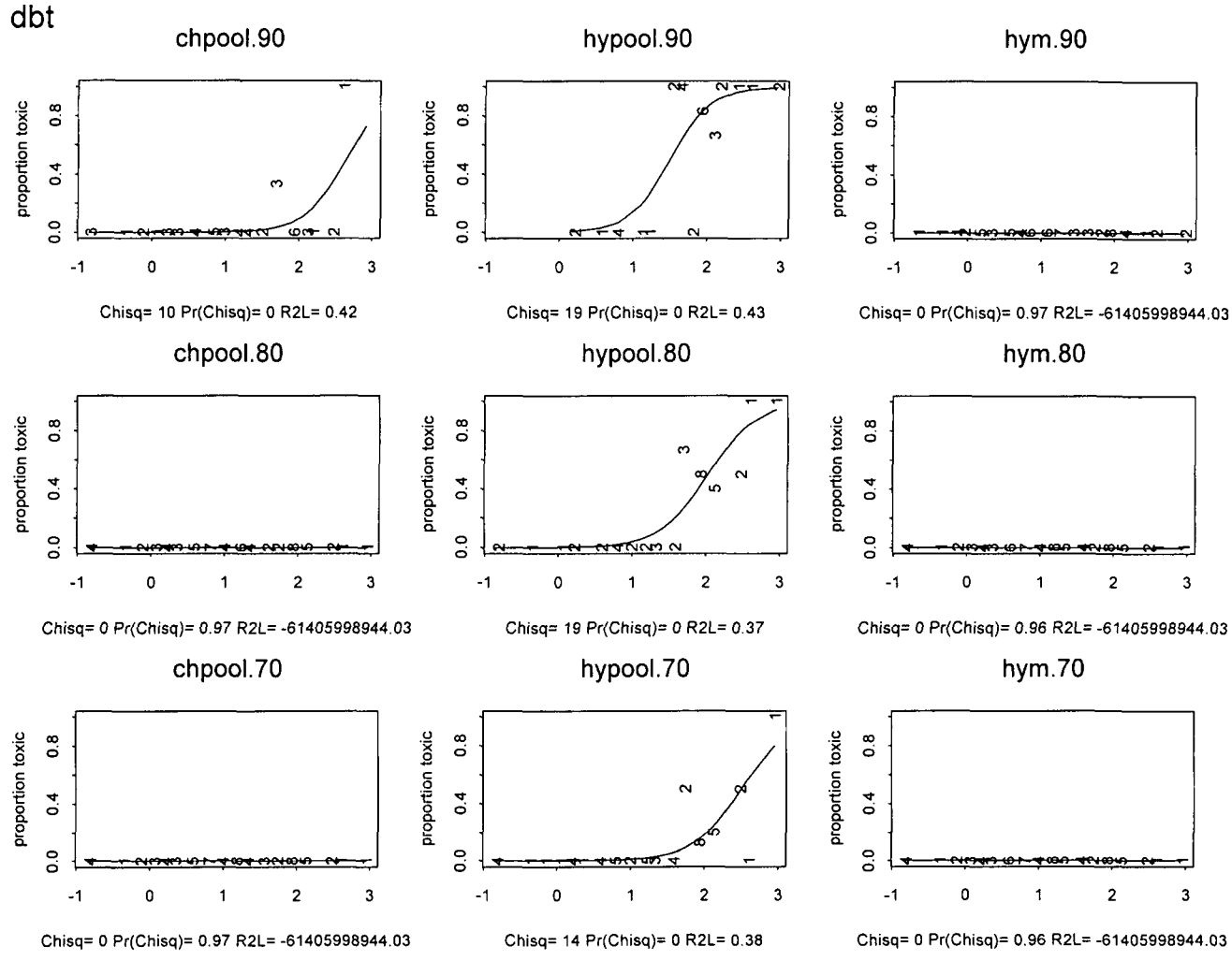
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-15. Logistic regression model – zinc



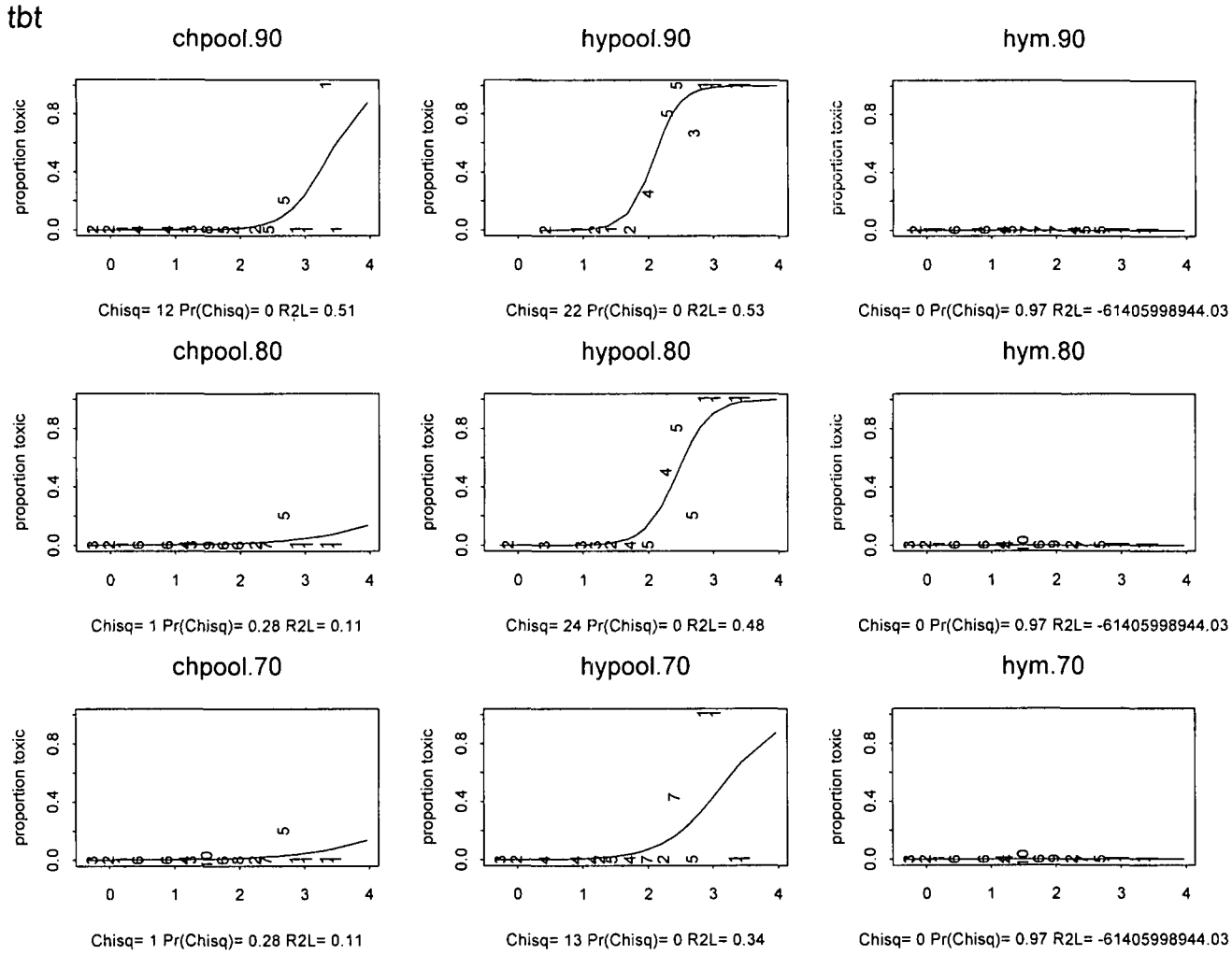
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-16. Logistic regression model – butyltin



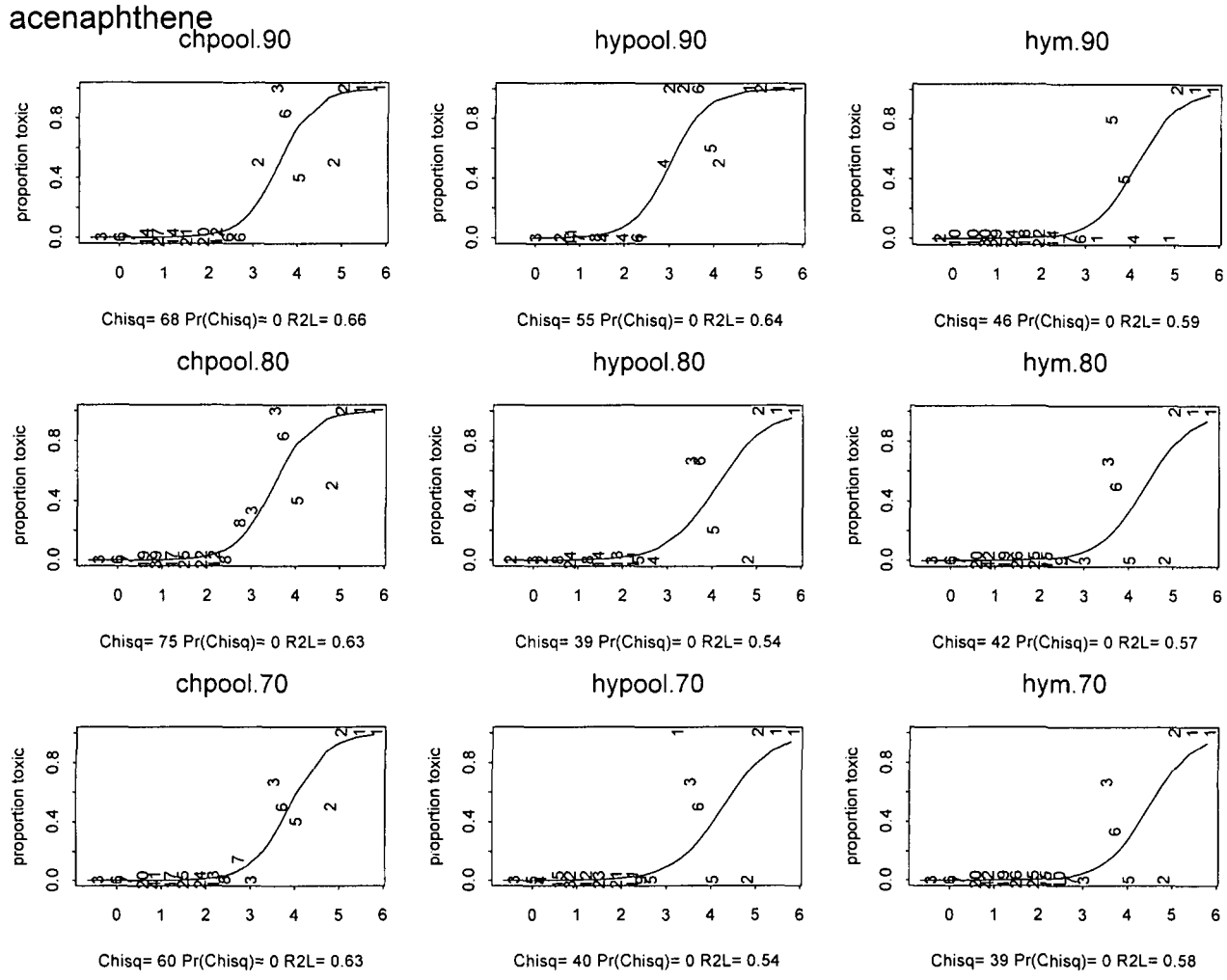
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-17. Logistic regression model – dibutyltin



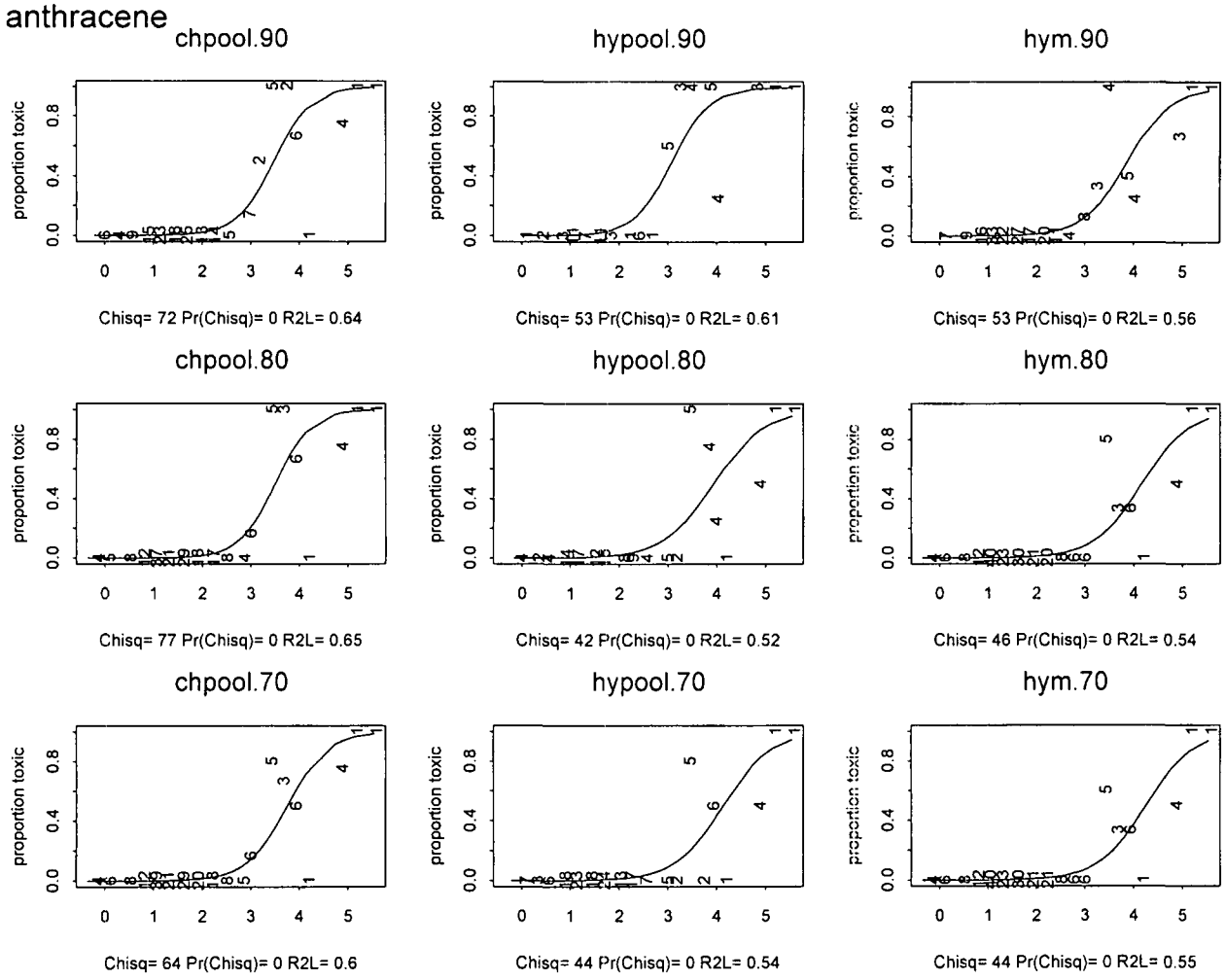
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-18. Logistic regression model – tributyltin



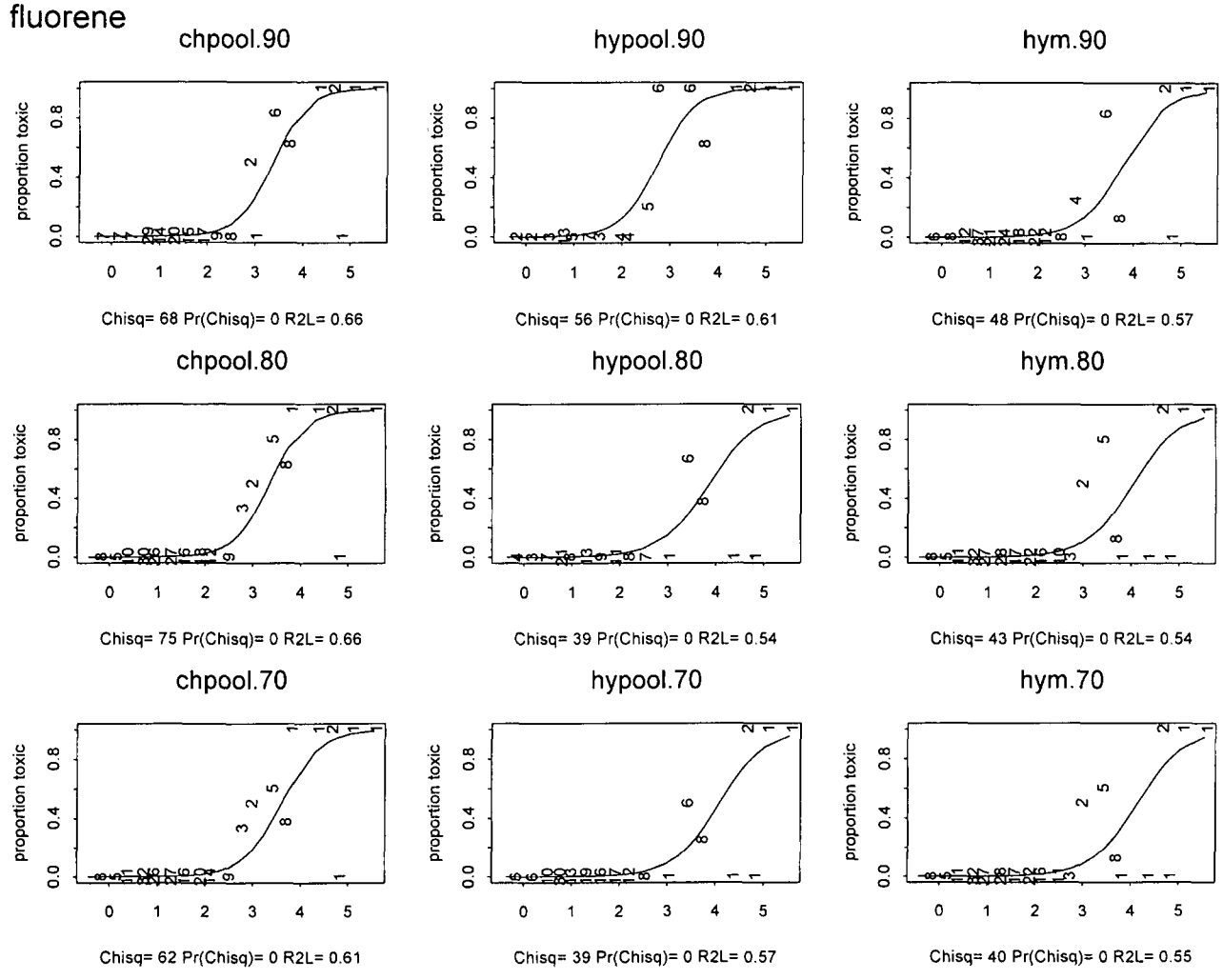
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-19. Logistic regression model – acenaphthene



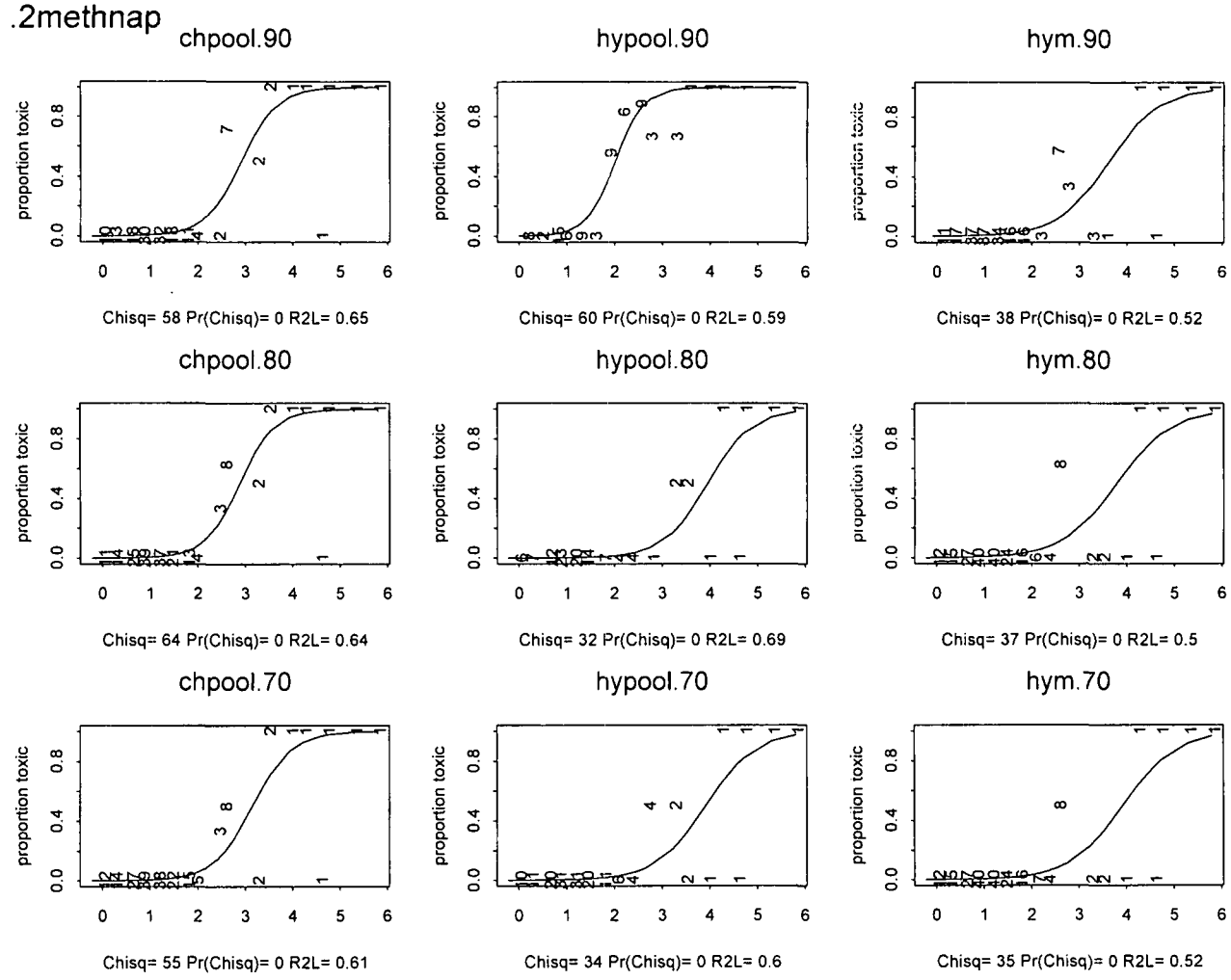
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-20. Logistic regression model – anthracene



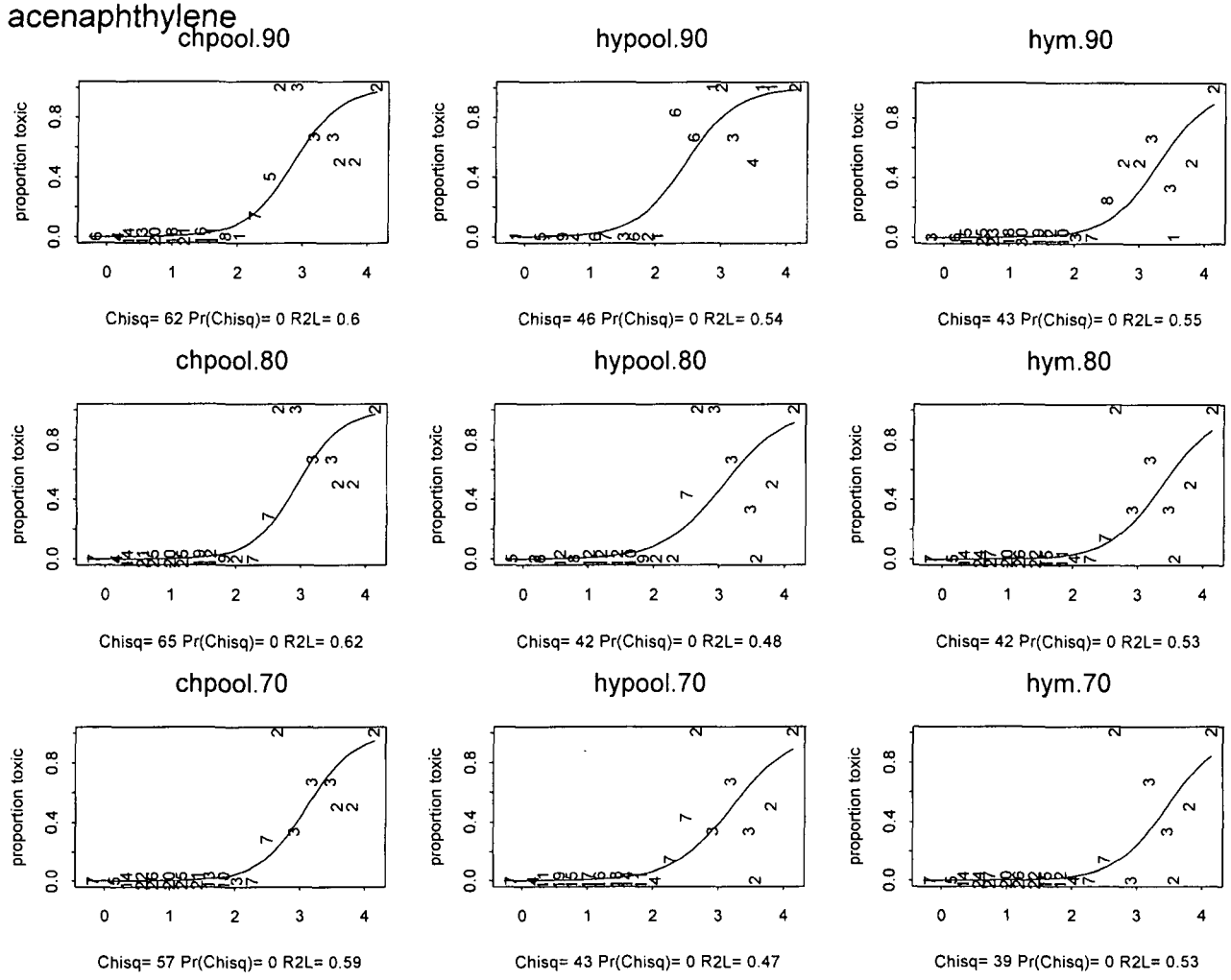
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-21. Logistic regression model – fluorene



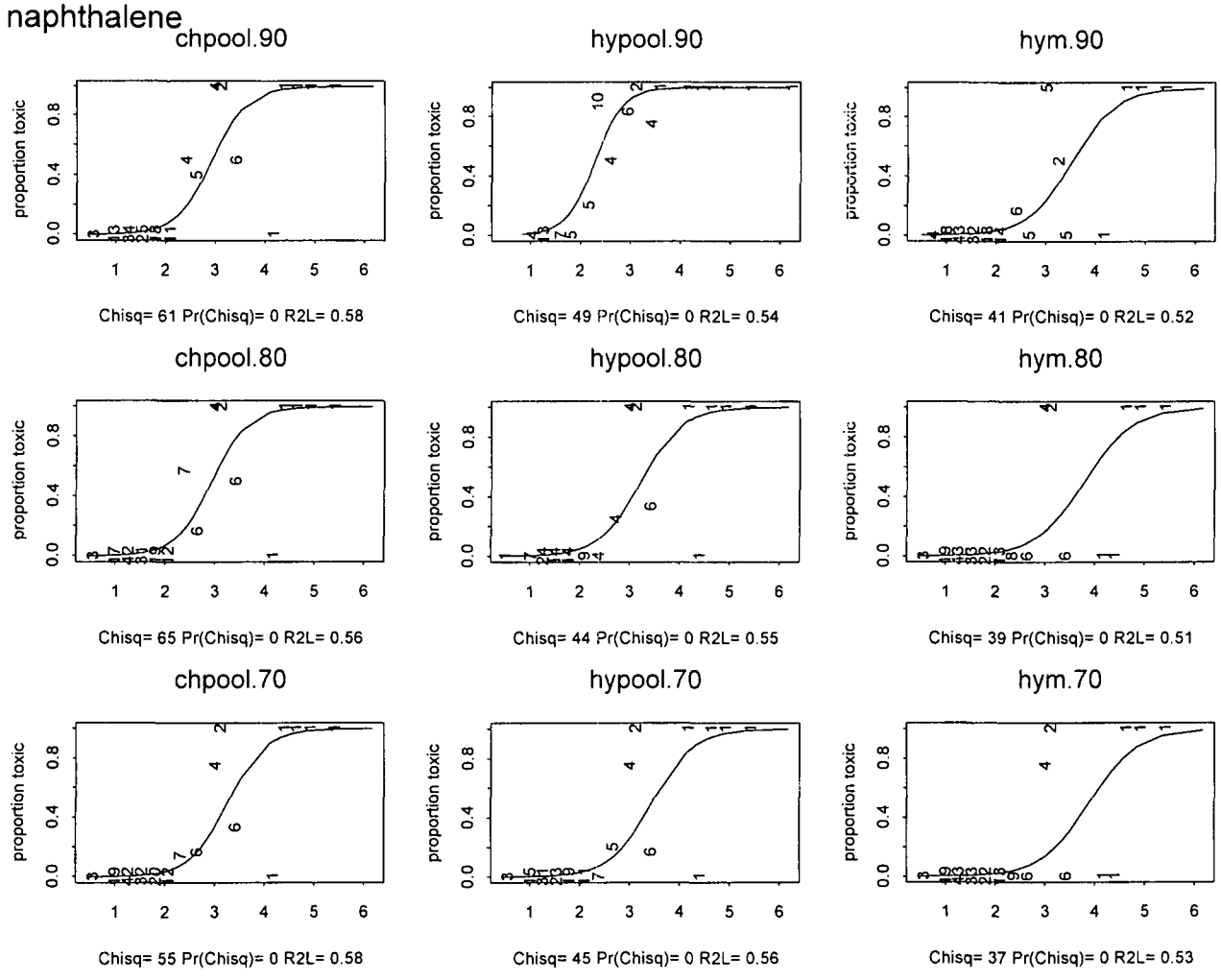
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-22. Logistic regression model – 2-methylnaphthalene



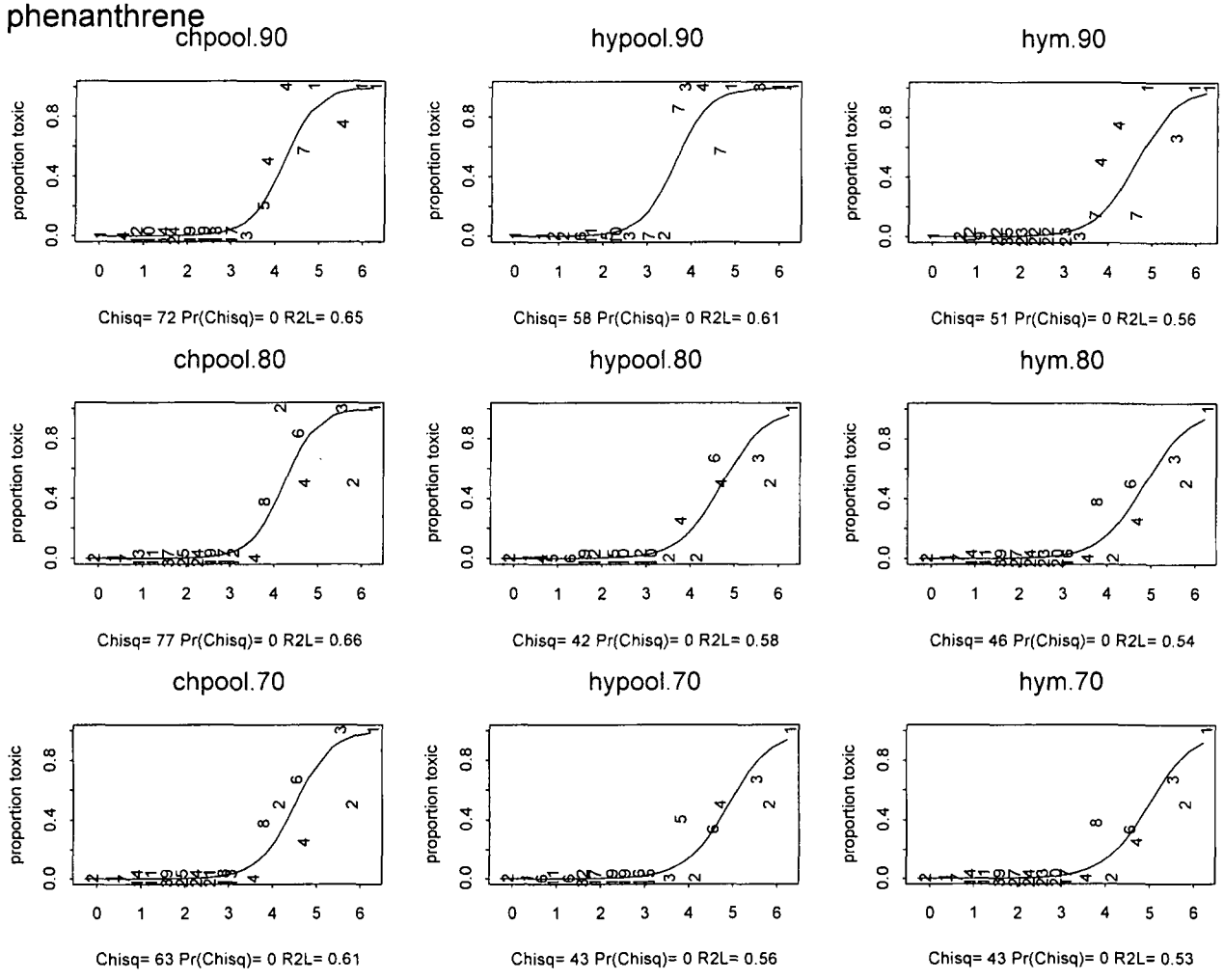
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-23. Logistic regression model – acenaphthylene



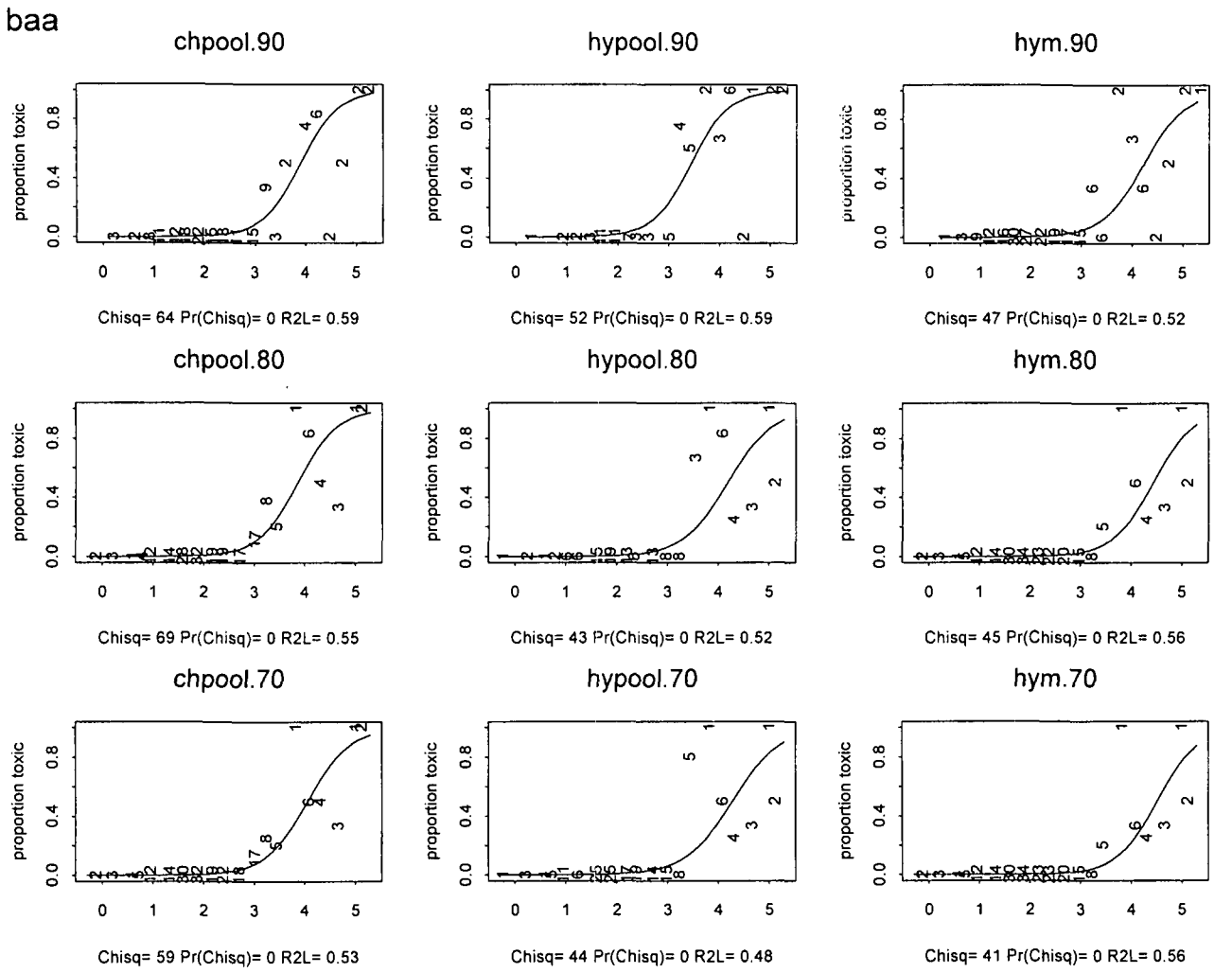
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-24. Logistic regression model – naphthalene



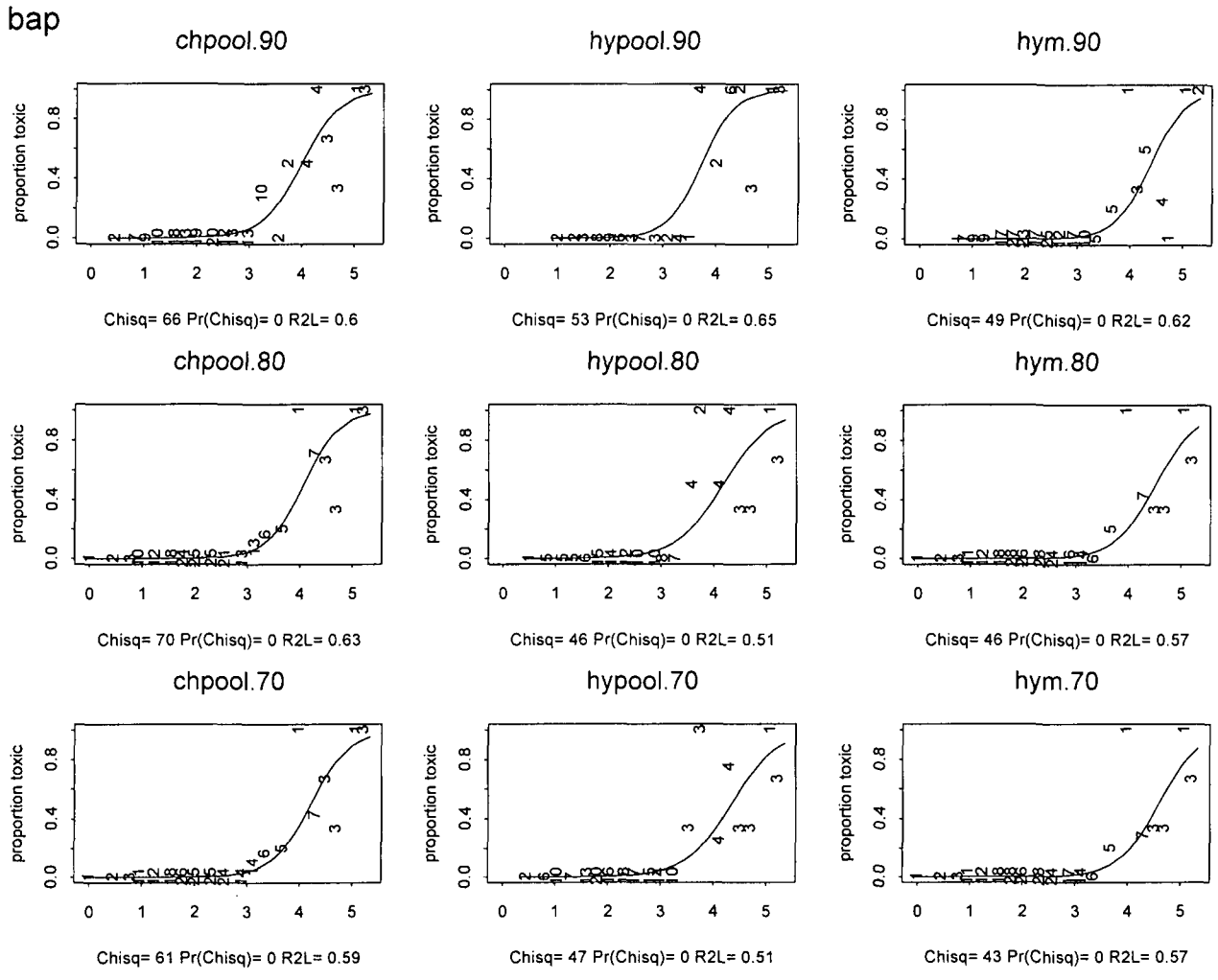
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-25. Logistic regression model – phenanthrene



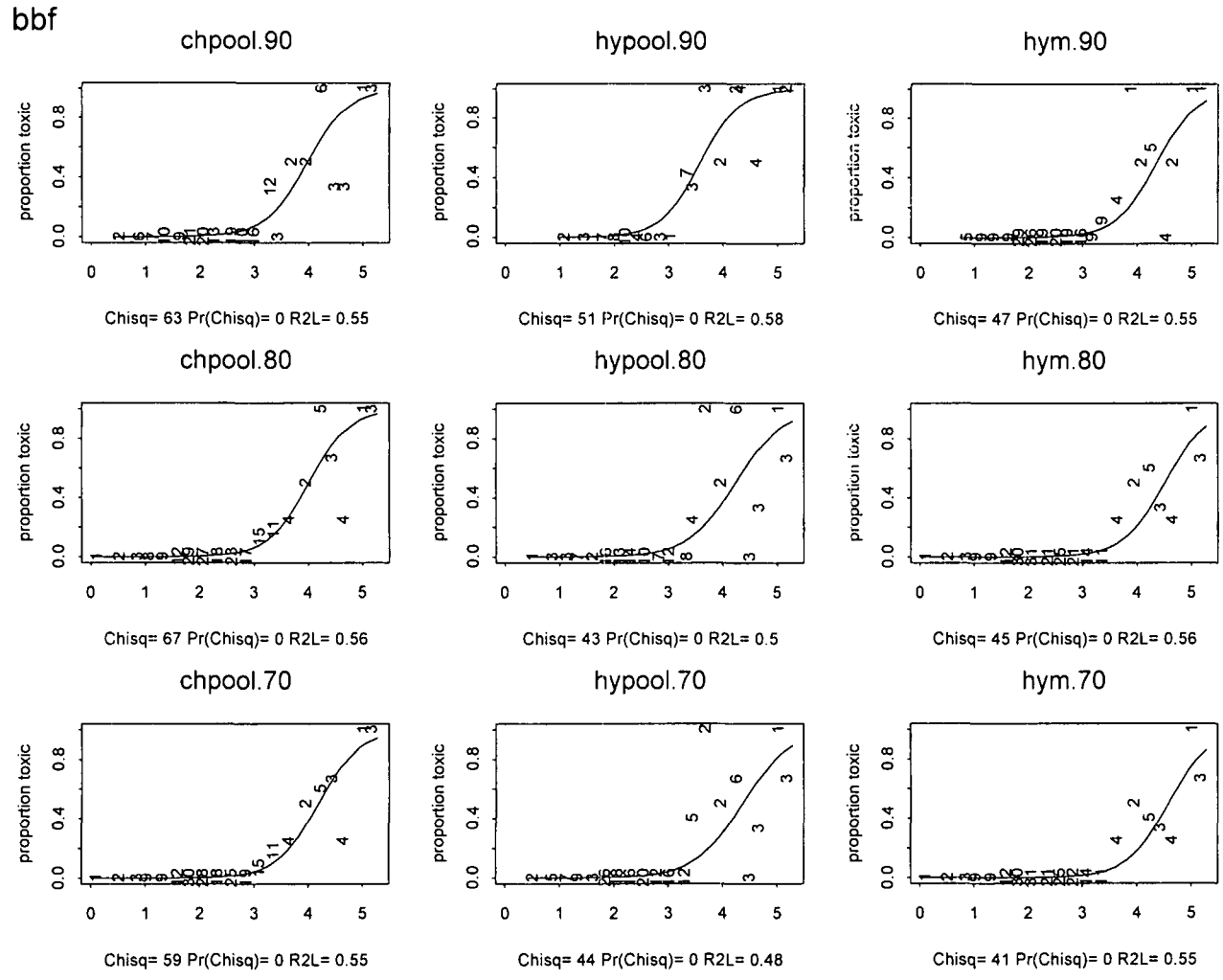
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-26. Logistic regression model – benzo(a)anthracene



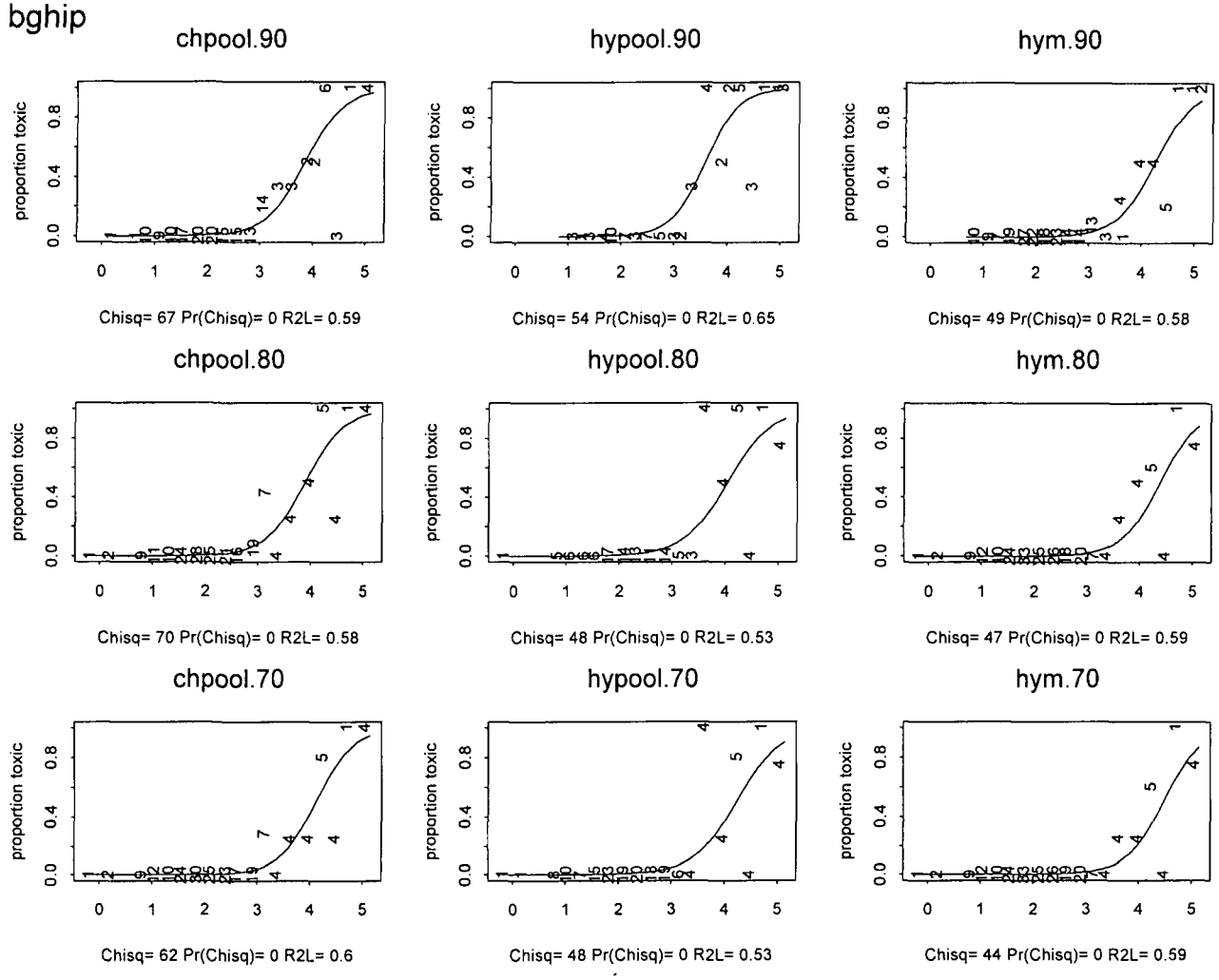
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-27. Logistic regression model – benzo(a)pyrene



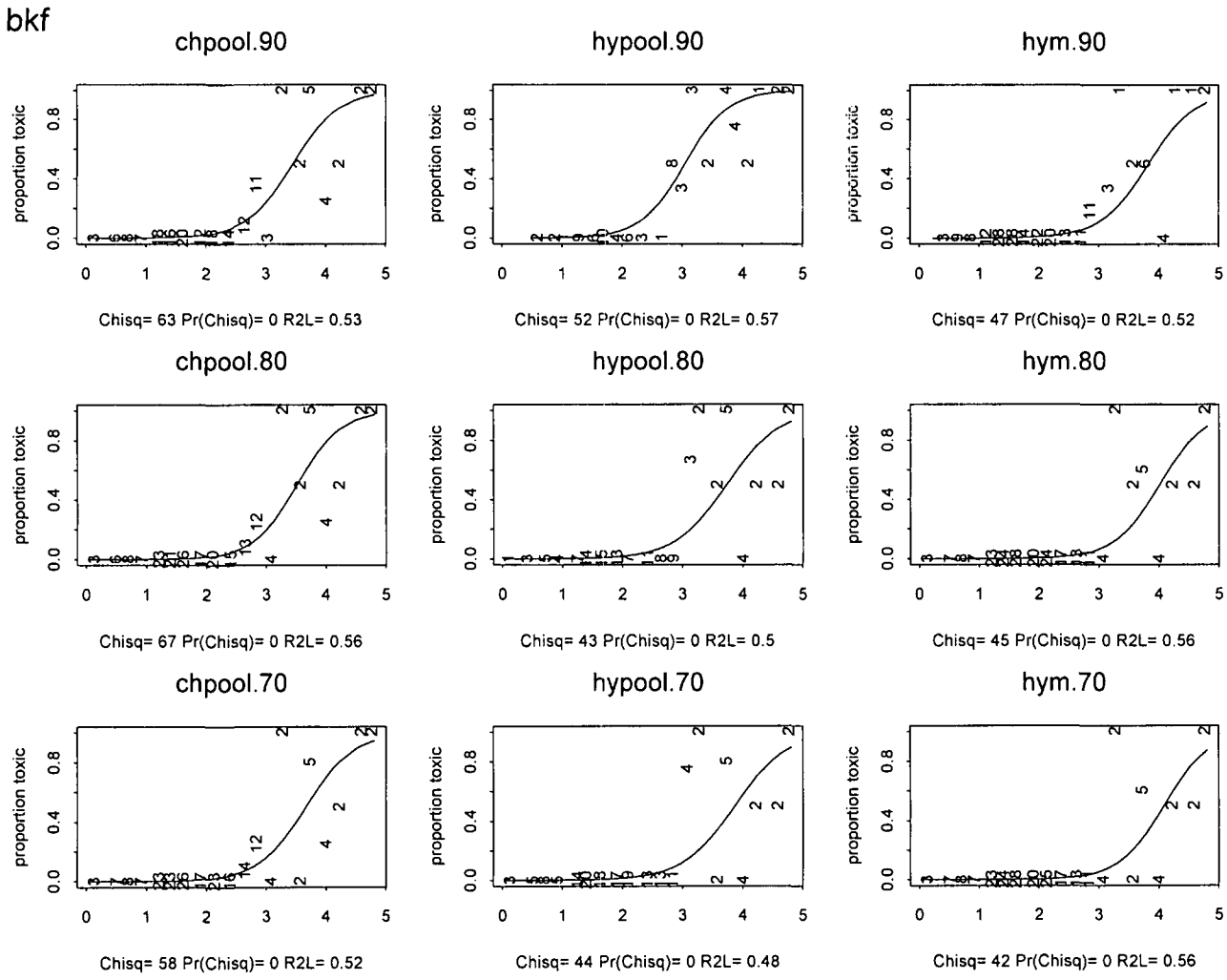
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-28. Logistic regression model – benzo(b)fluoranthene



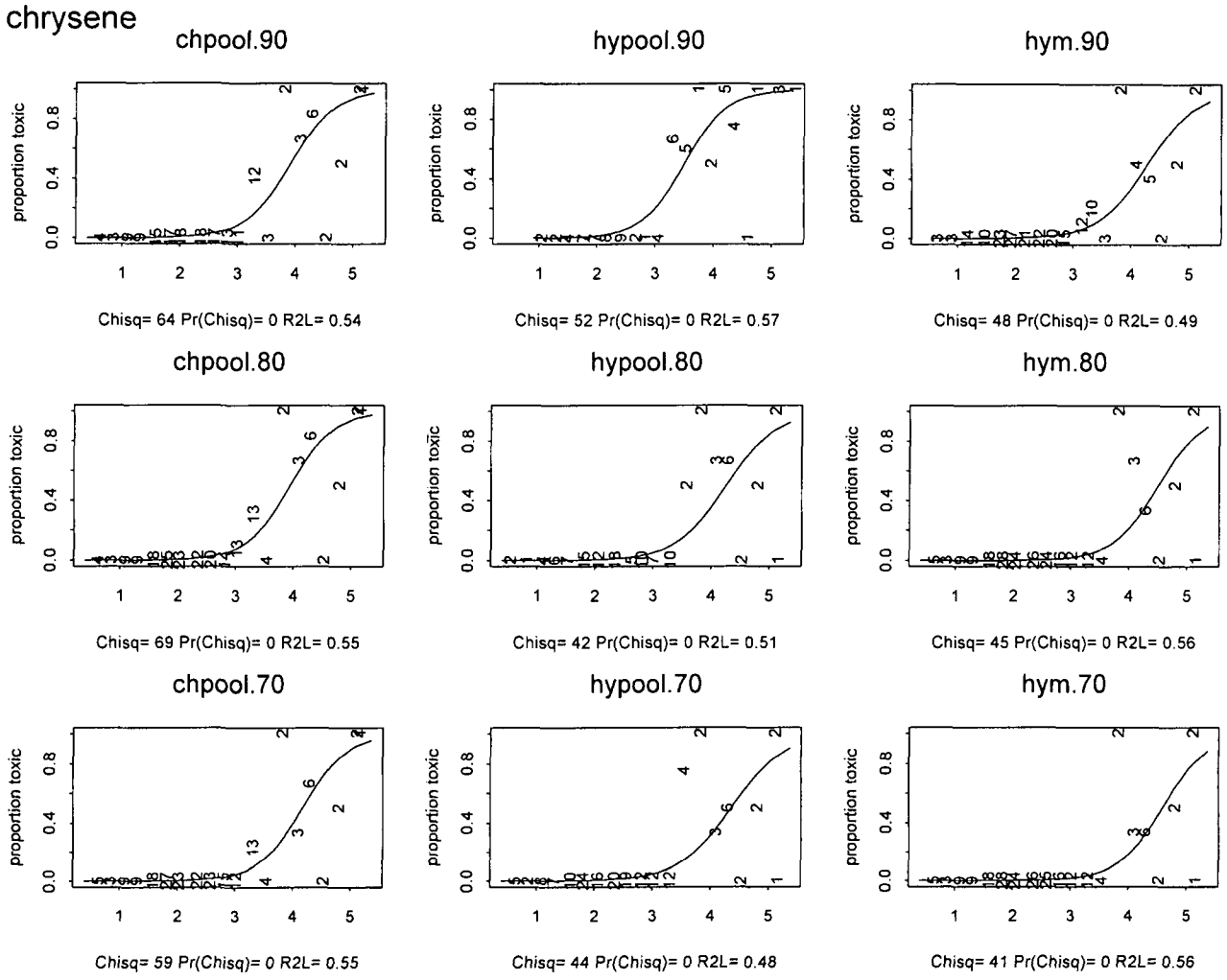
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-29. Logistic regression model – benzo(ghi)perylene



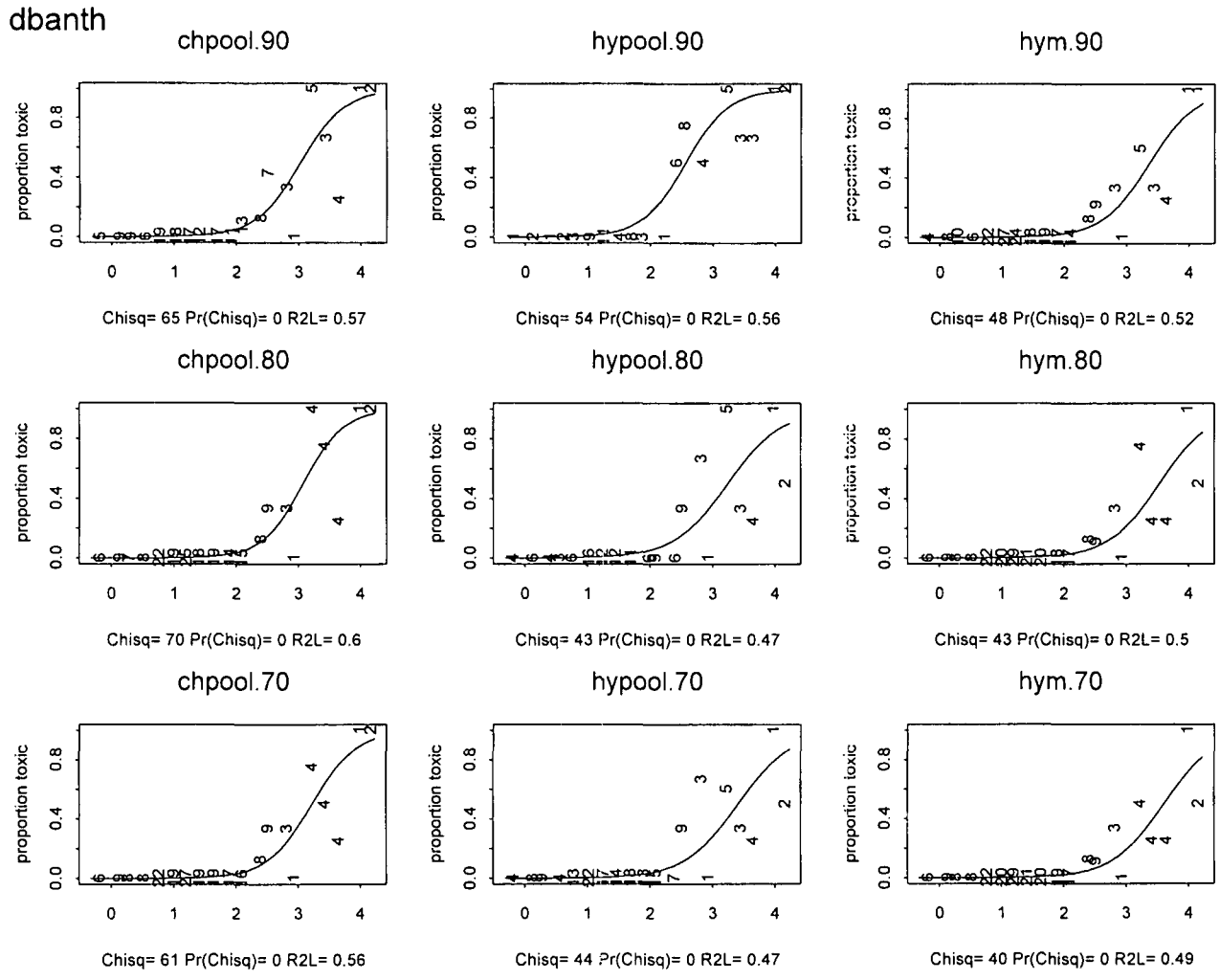
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-30. Logistic regression model – benzo(k)fluoranthene



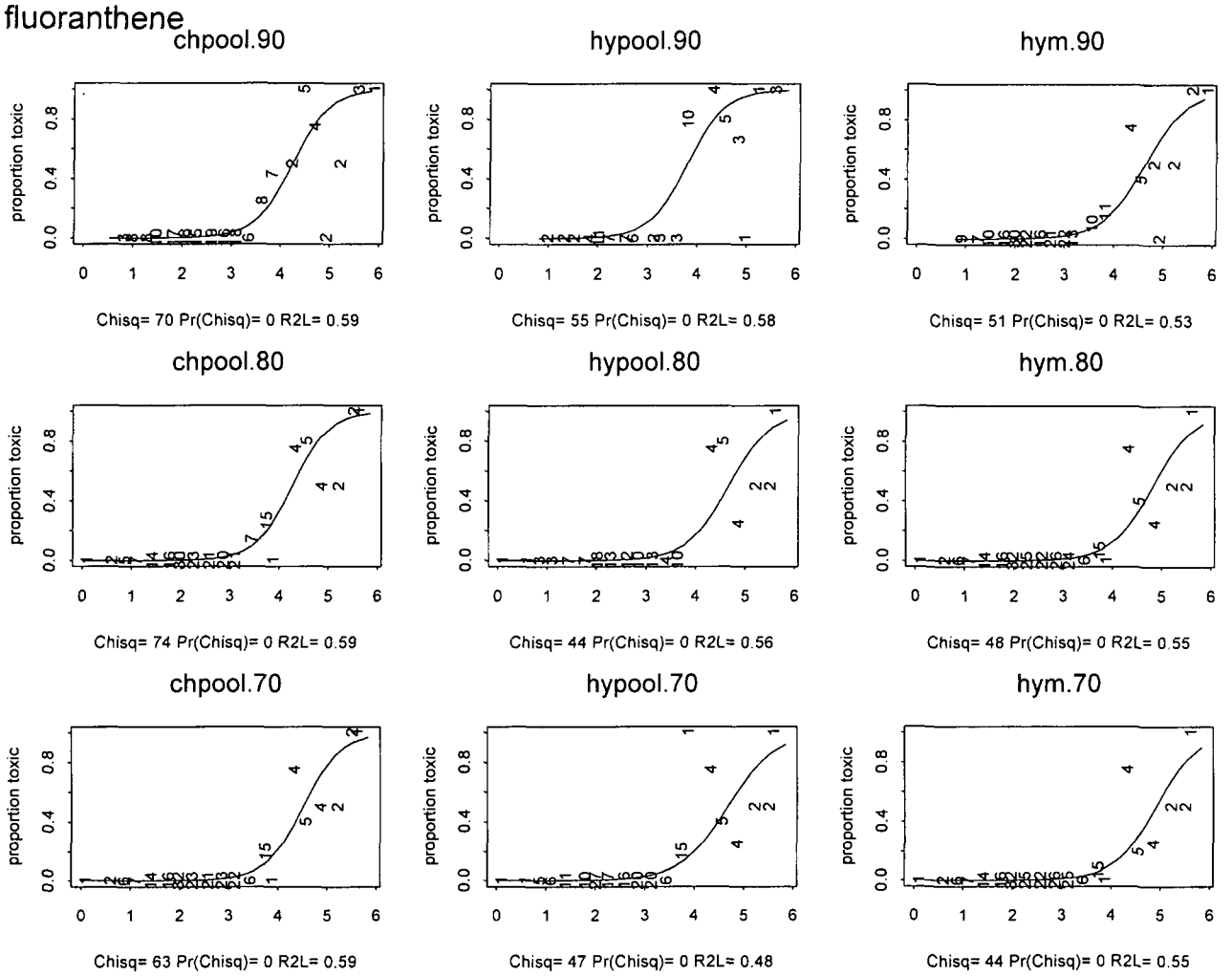
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-31. Logistic regression model – chrysene



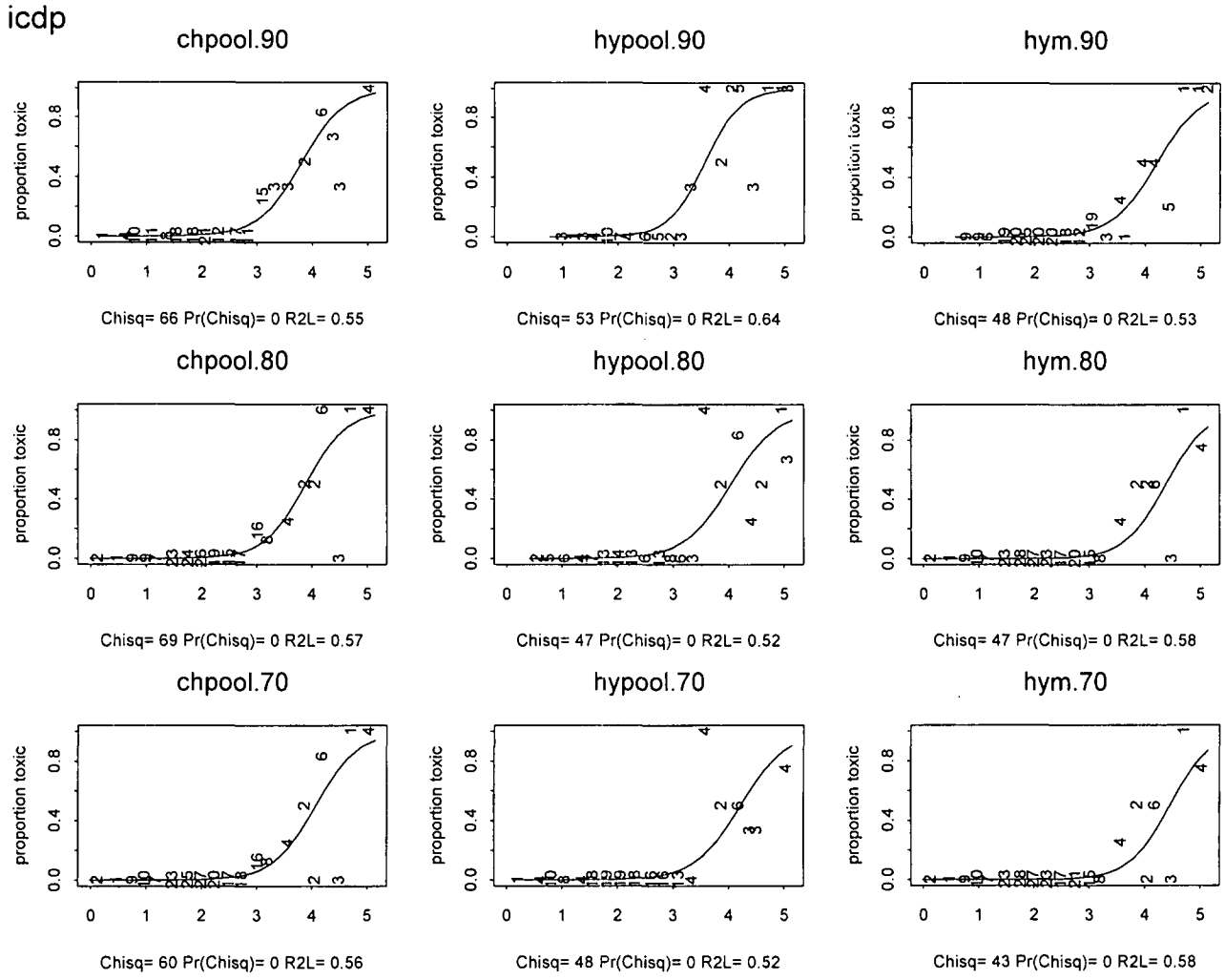
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-32. Logistic regression model – dibenzanthracene



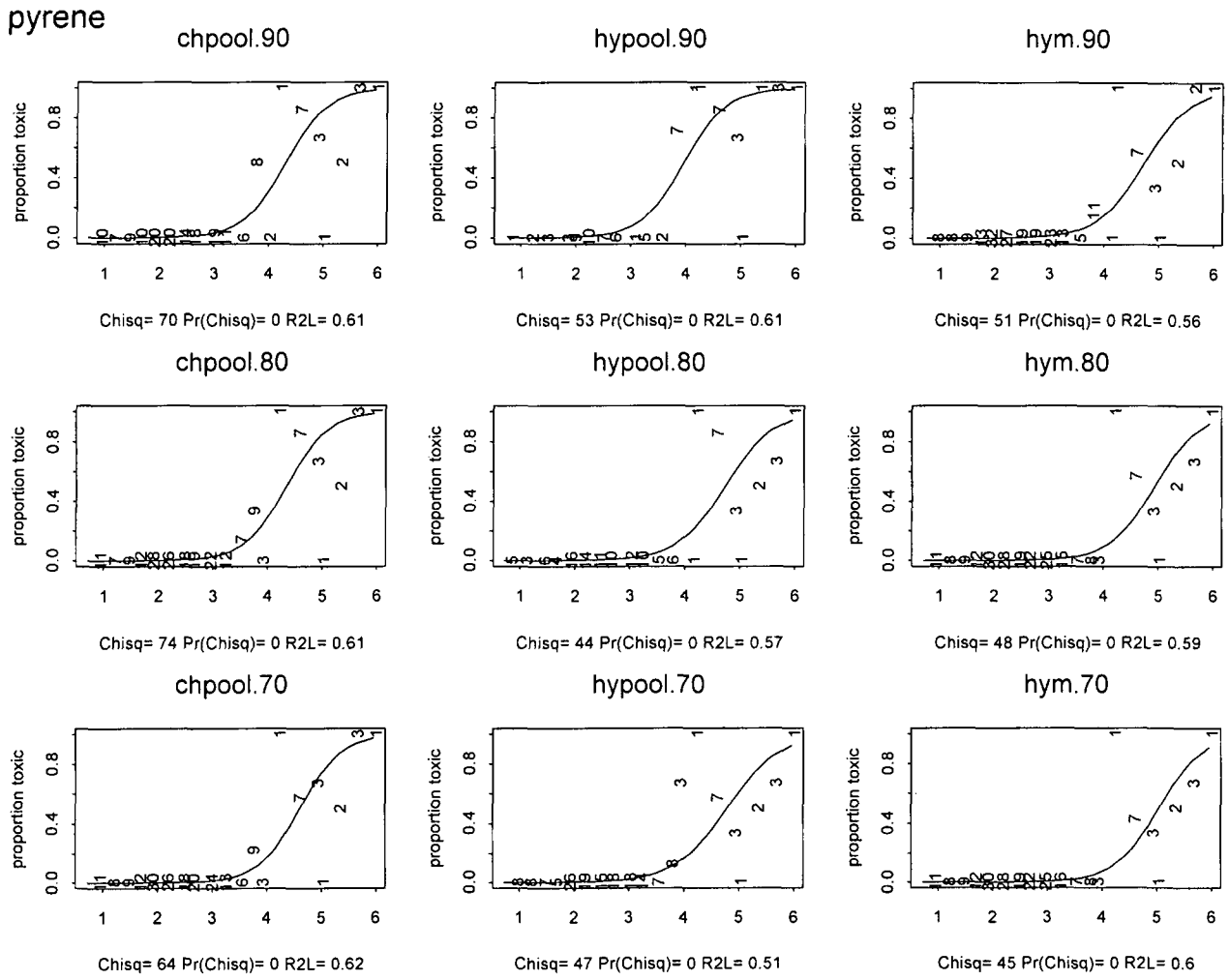
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-33. Logistic regression model – fluoranthene



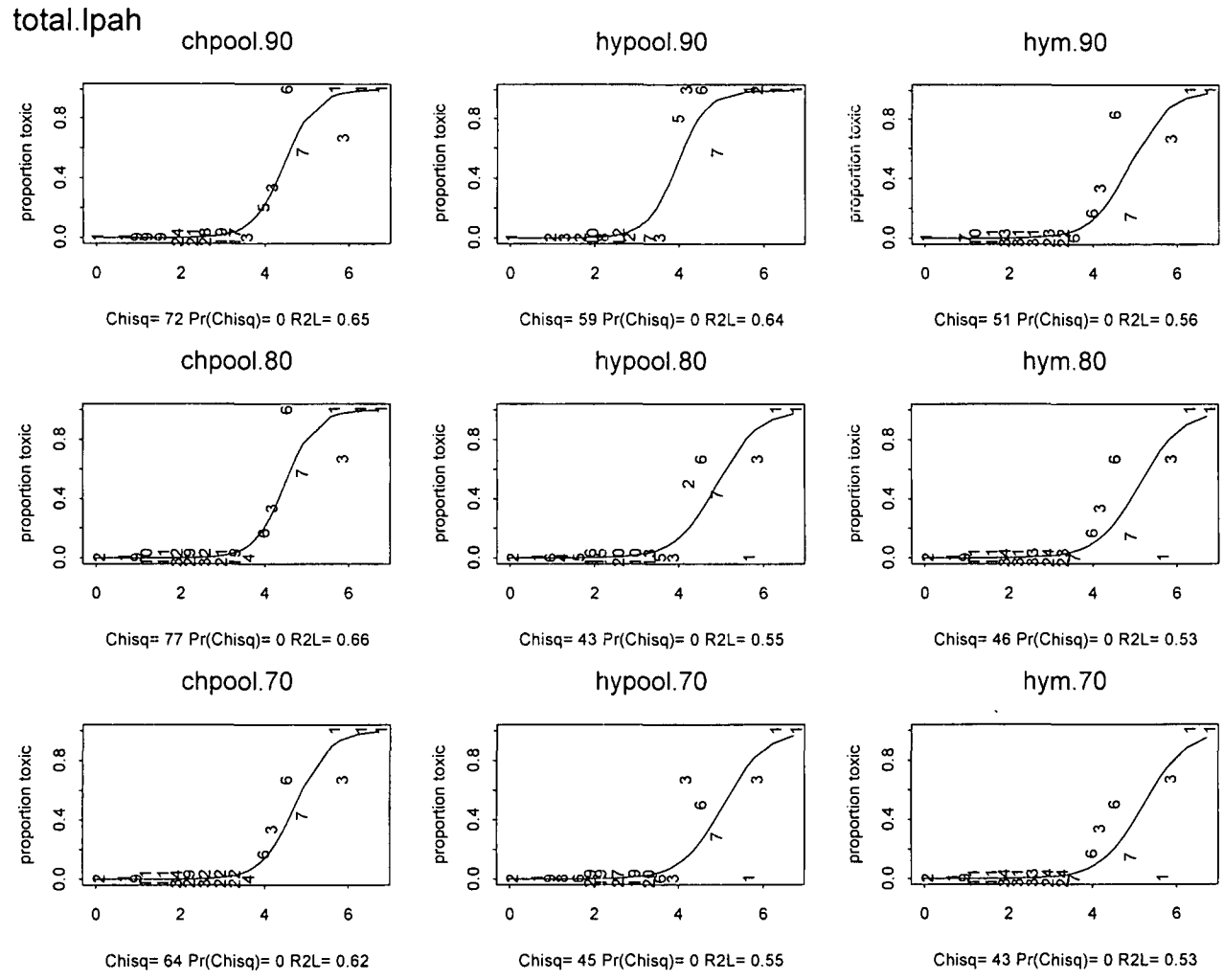
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-34. Logistic regression model – indeno(c,d)pyrene



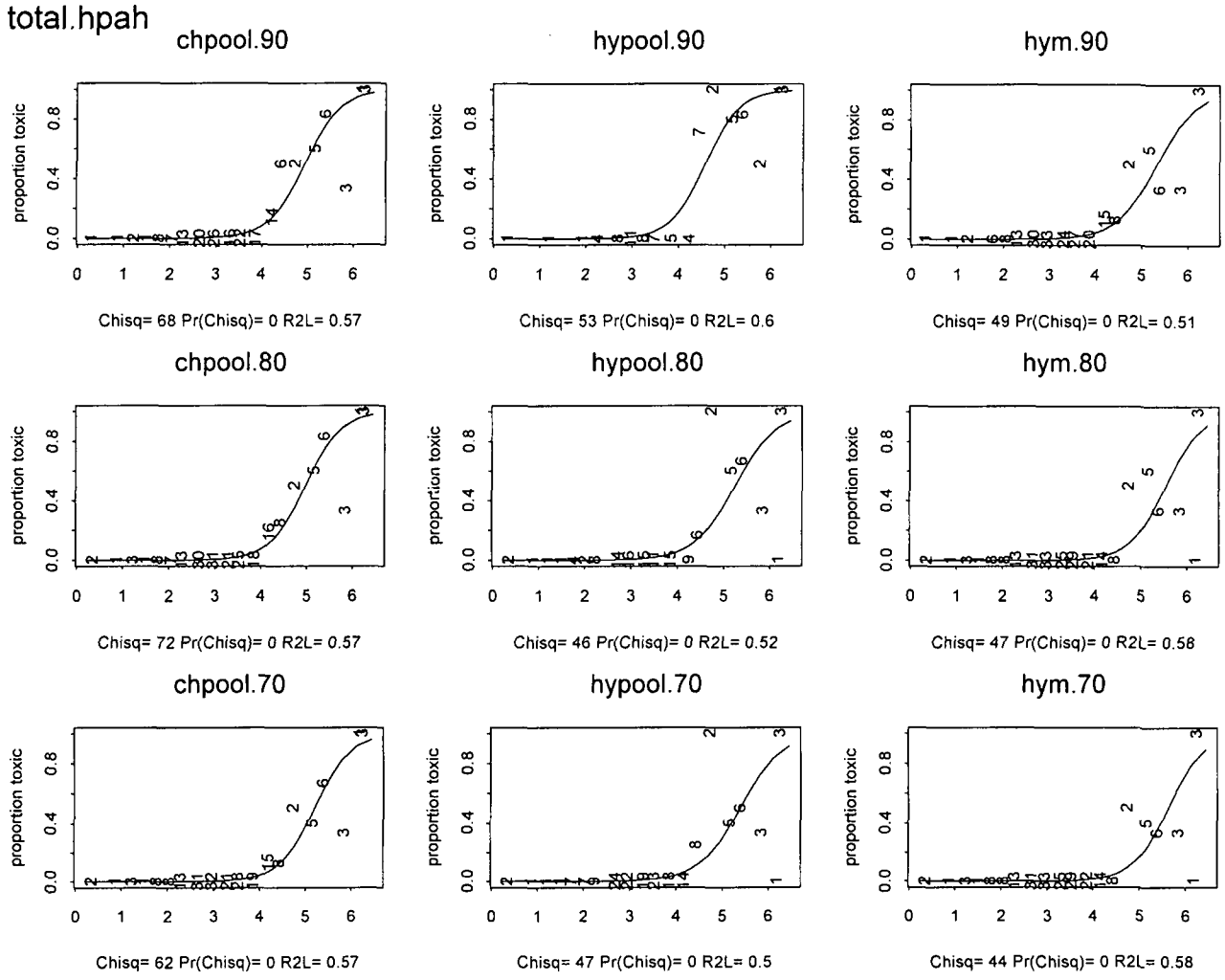
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-35. Logistic regression model – pyrene



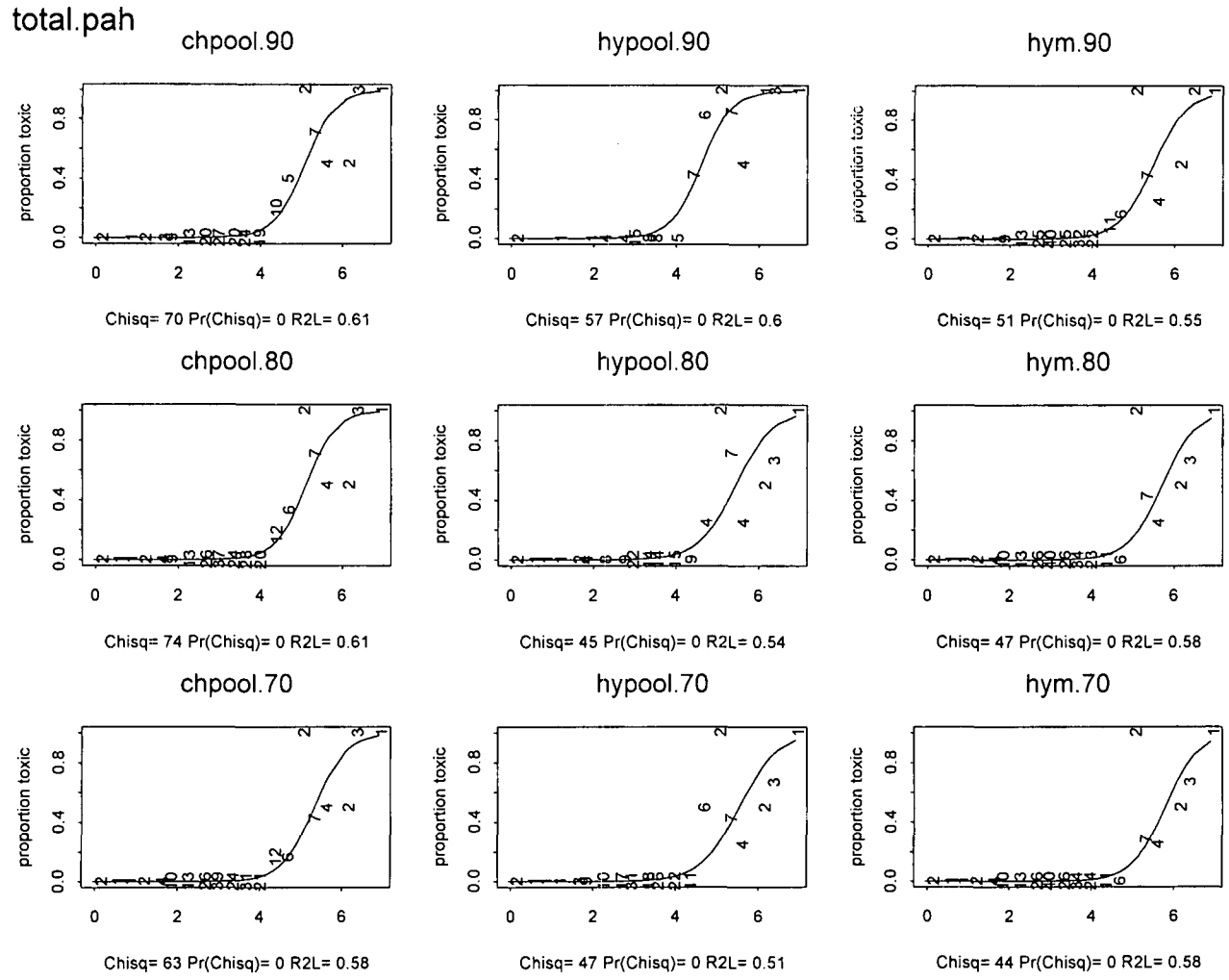
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-36. Logistic regression model – total LPAH



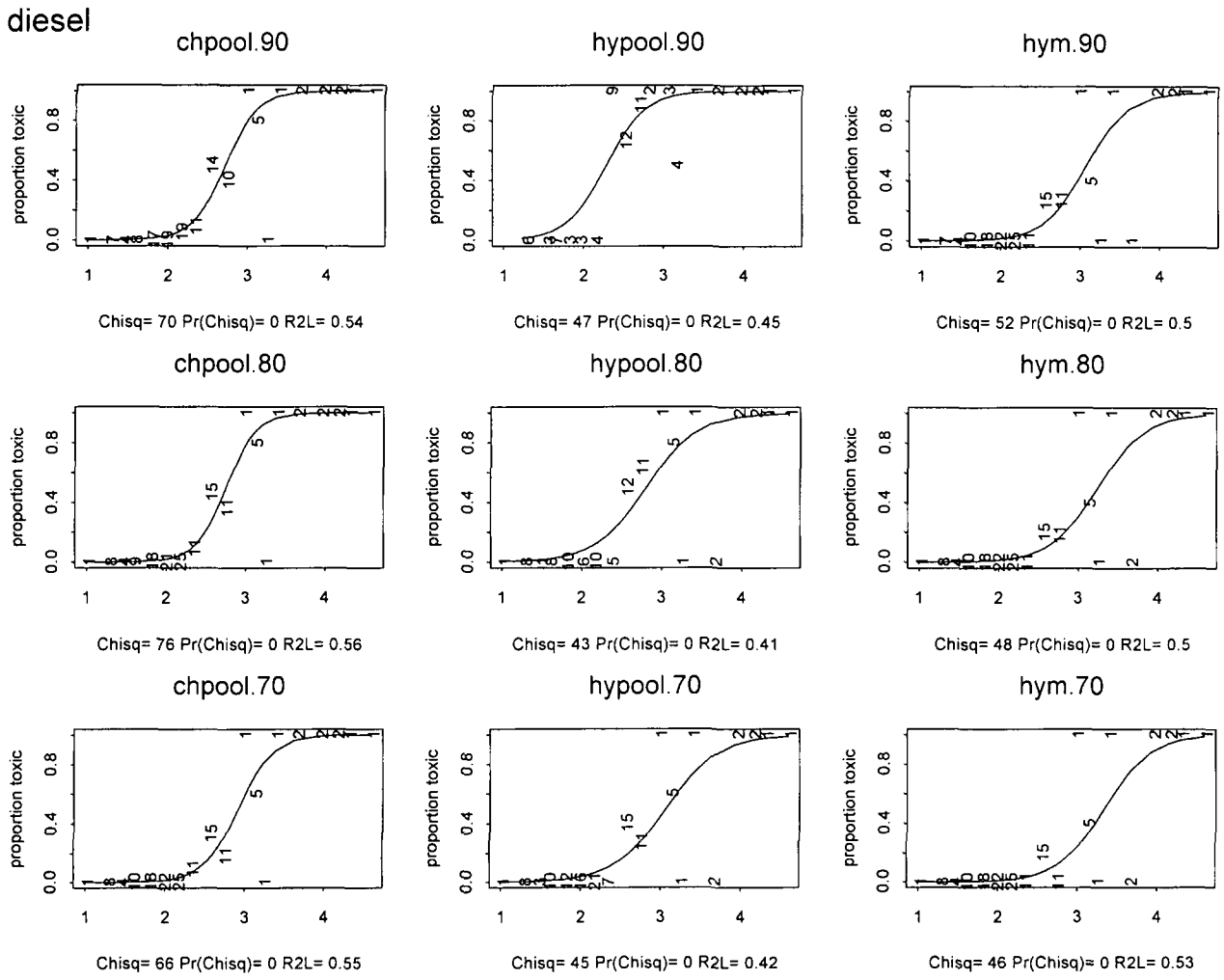
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-37. Logistic regression model – total HPAH



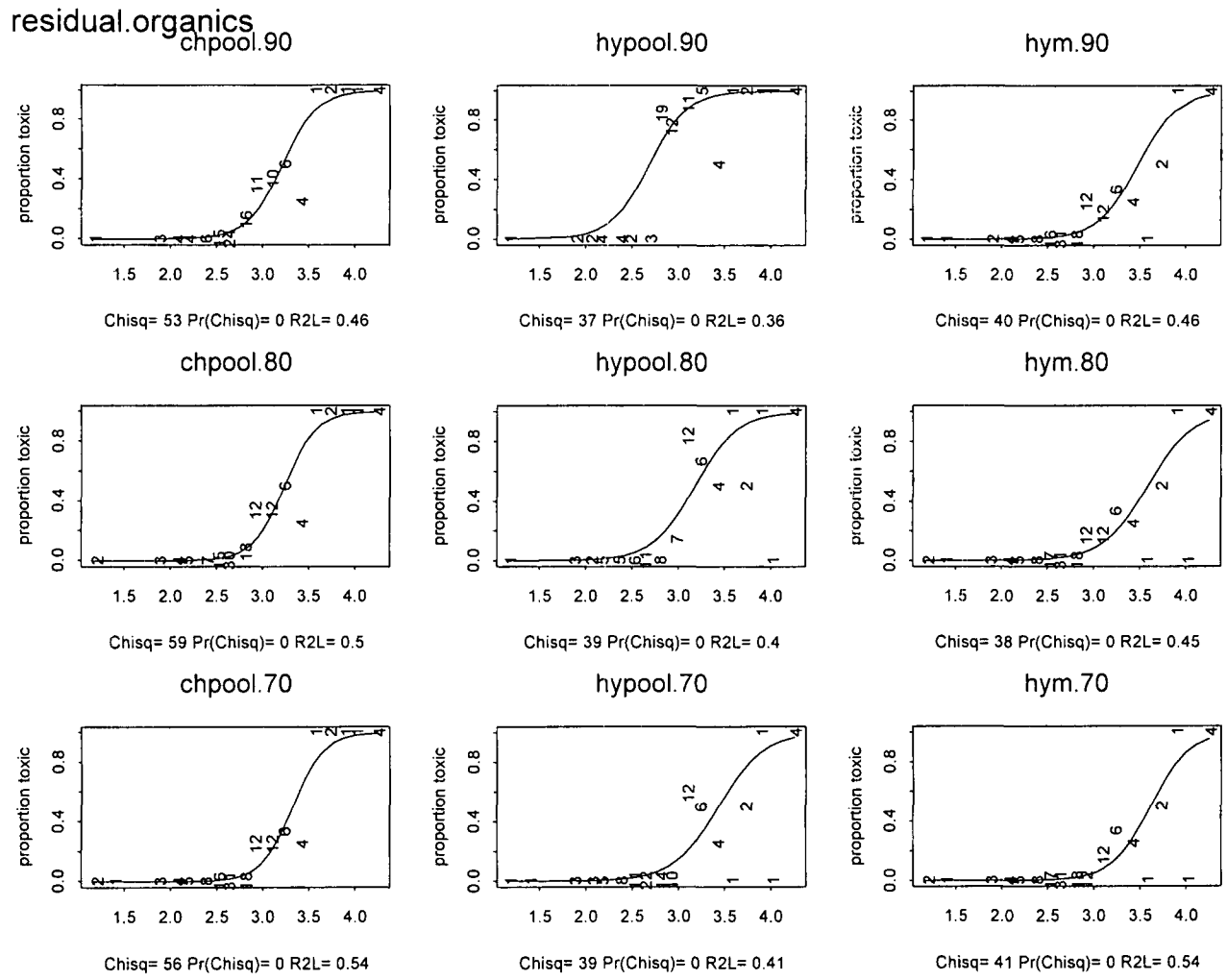
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-38. Logistic regression model – total PAH



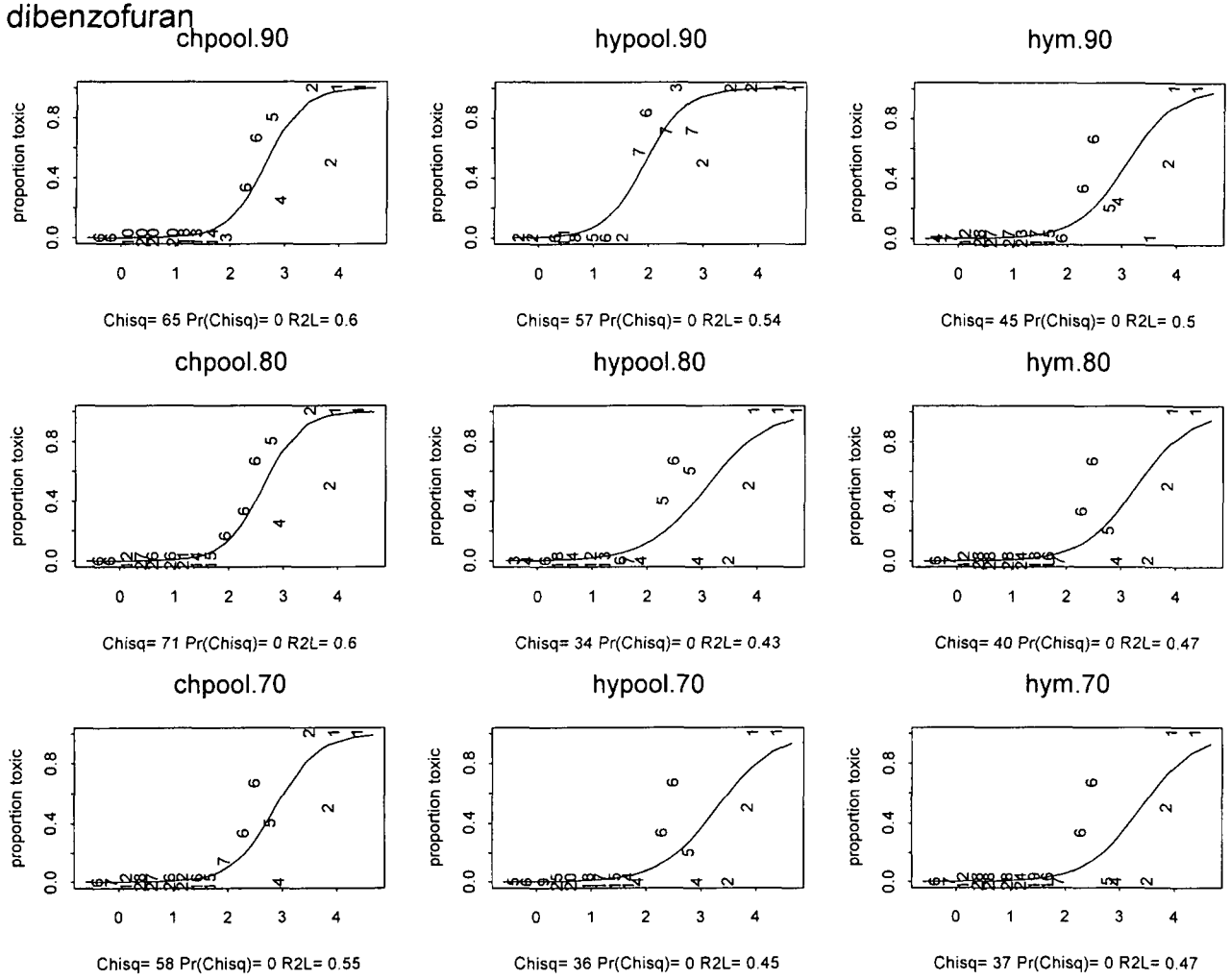
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-39. Logistic regression model – diesel-range hydrocarbons



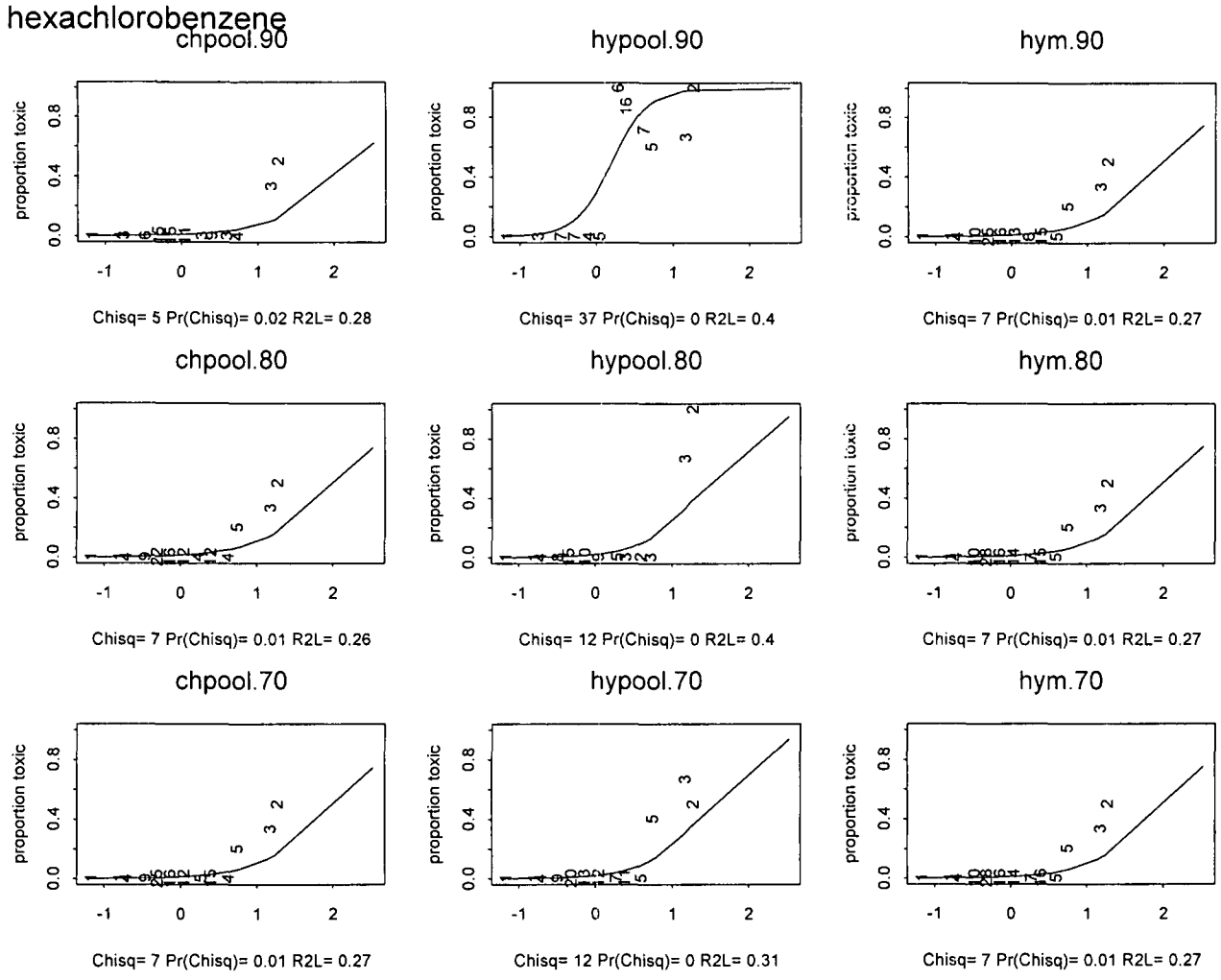
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-40. Logistic regression model – residual-range hydrocarbons



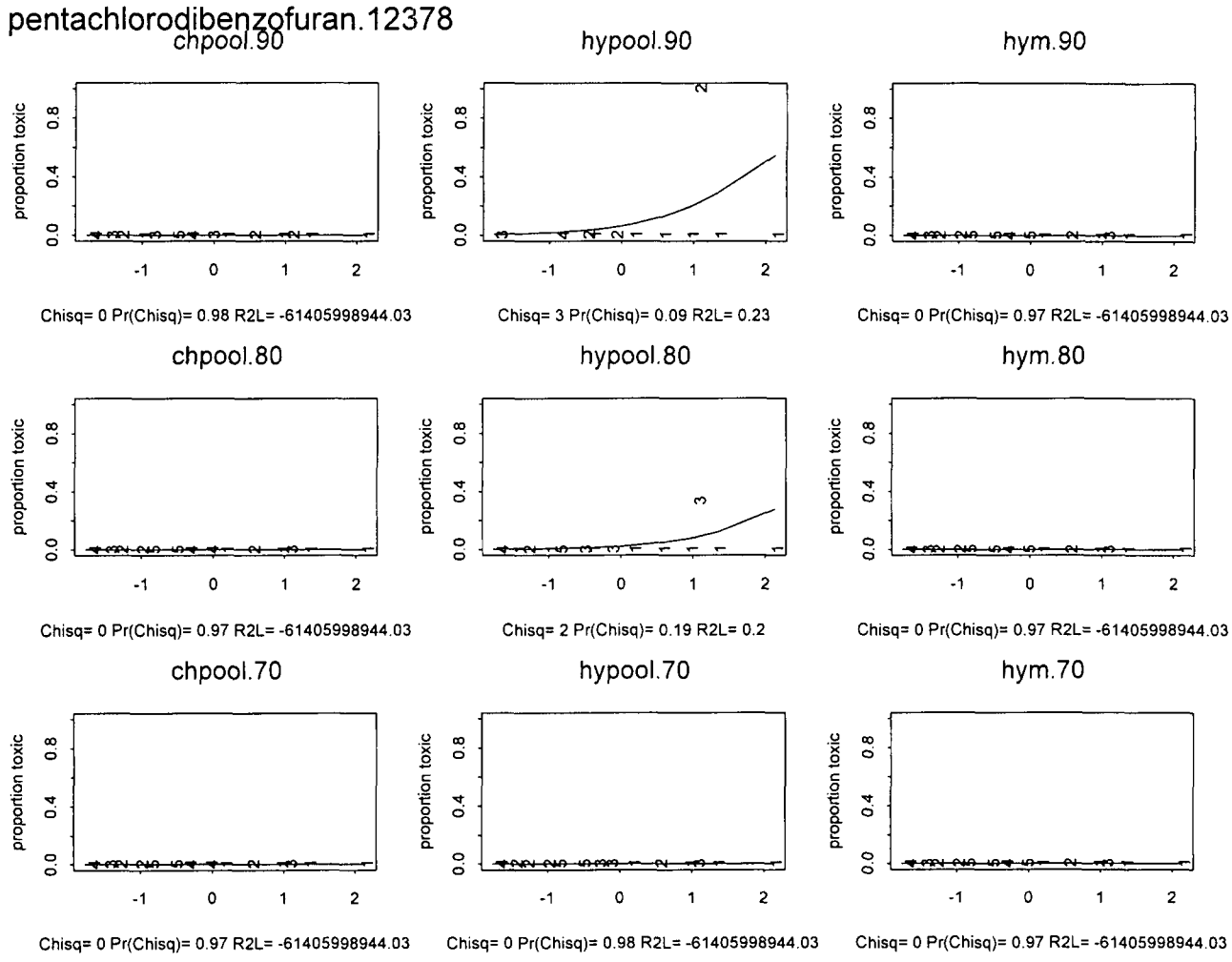
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-41. Logistic regression model – dibenzofuran



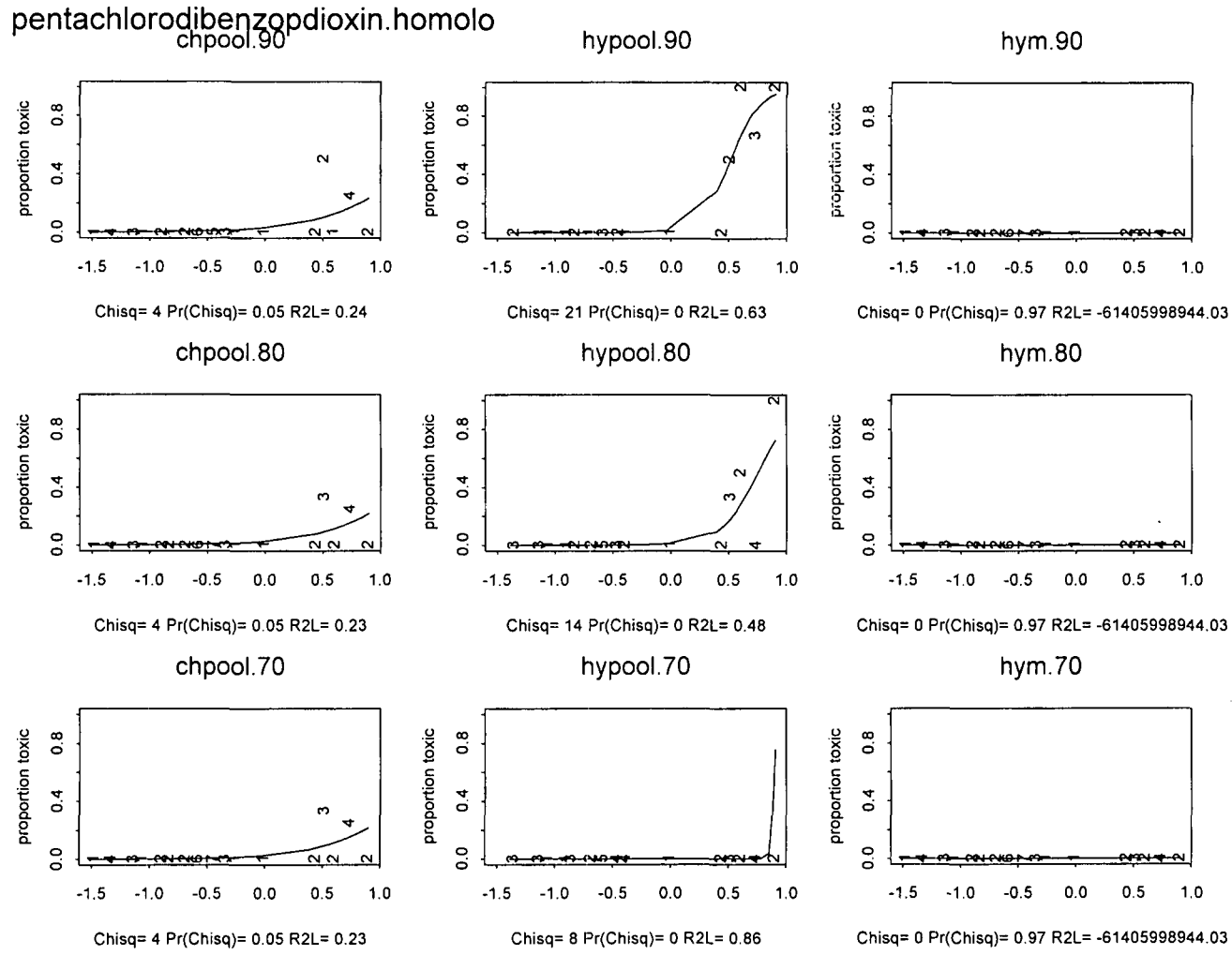
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-42. Logistic regression model – hexachlorobenzene



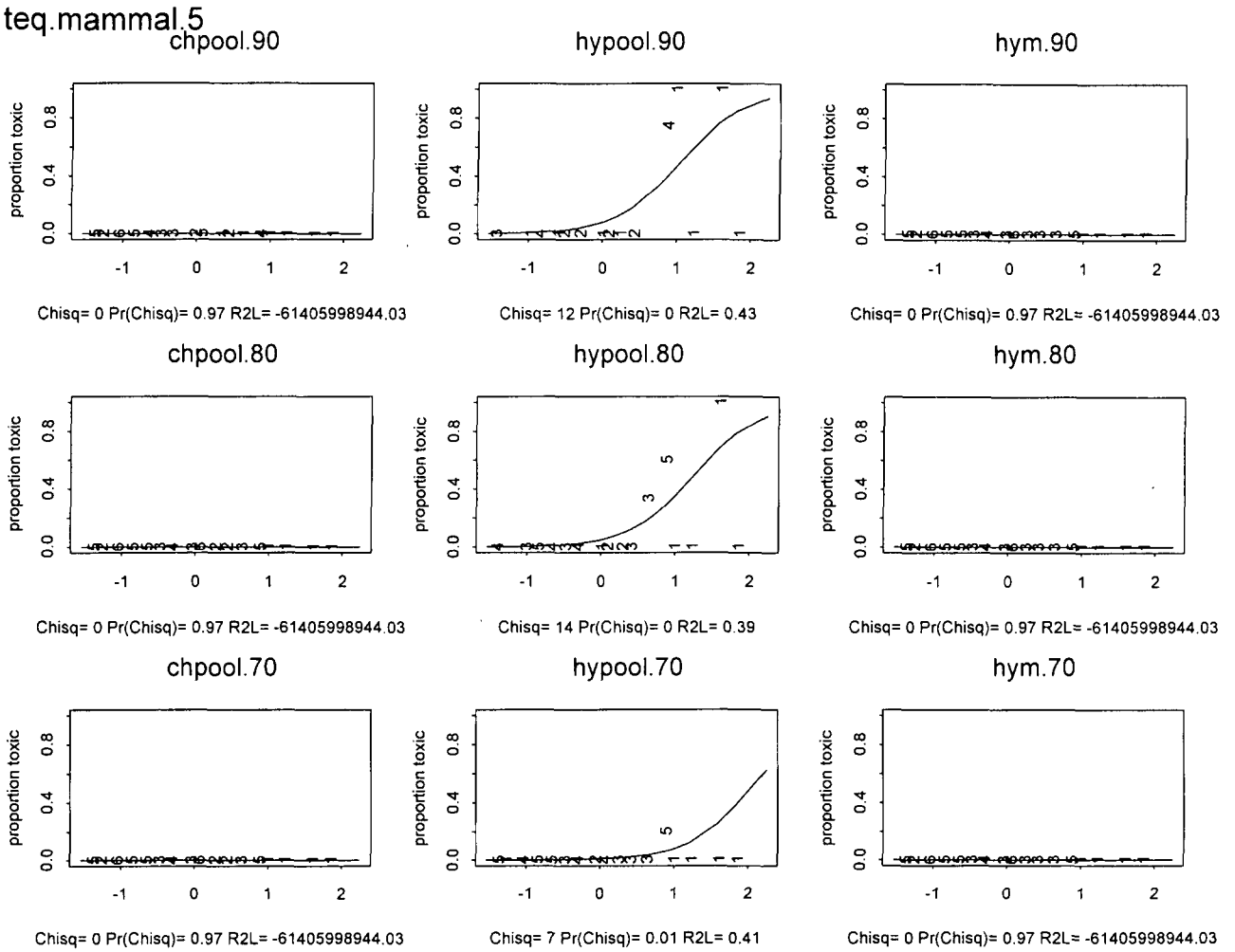
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-43. Logistic regression model – 1,2,3,7,8-pentachlorodibenzofuran



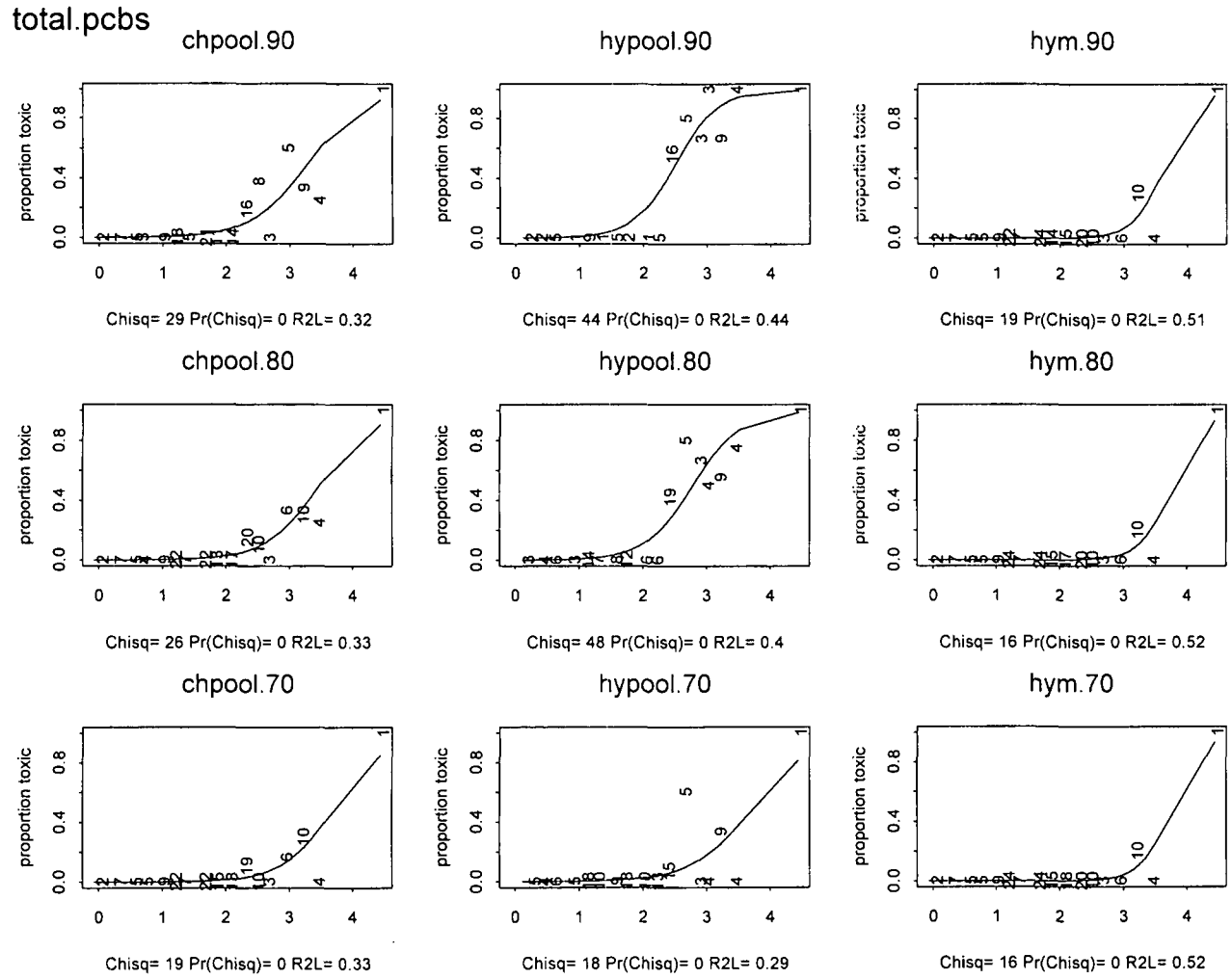
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-44. Logistic regression model – pentachlorodibenzo-p-dioxin homologs



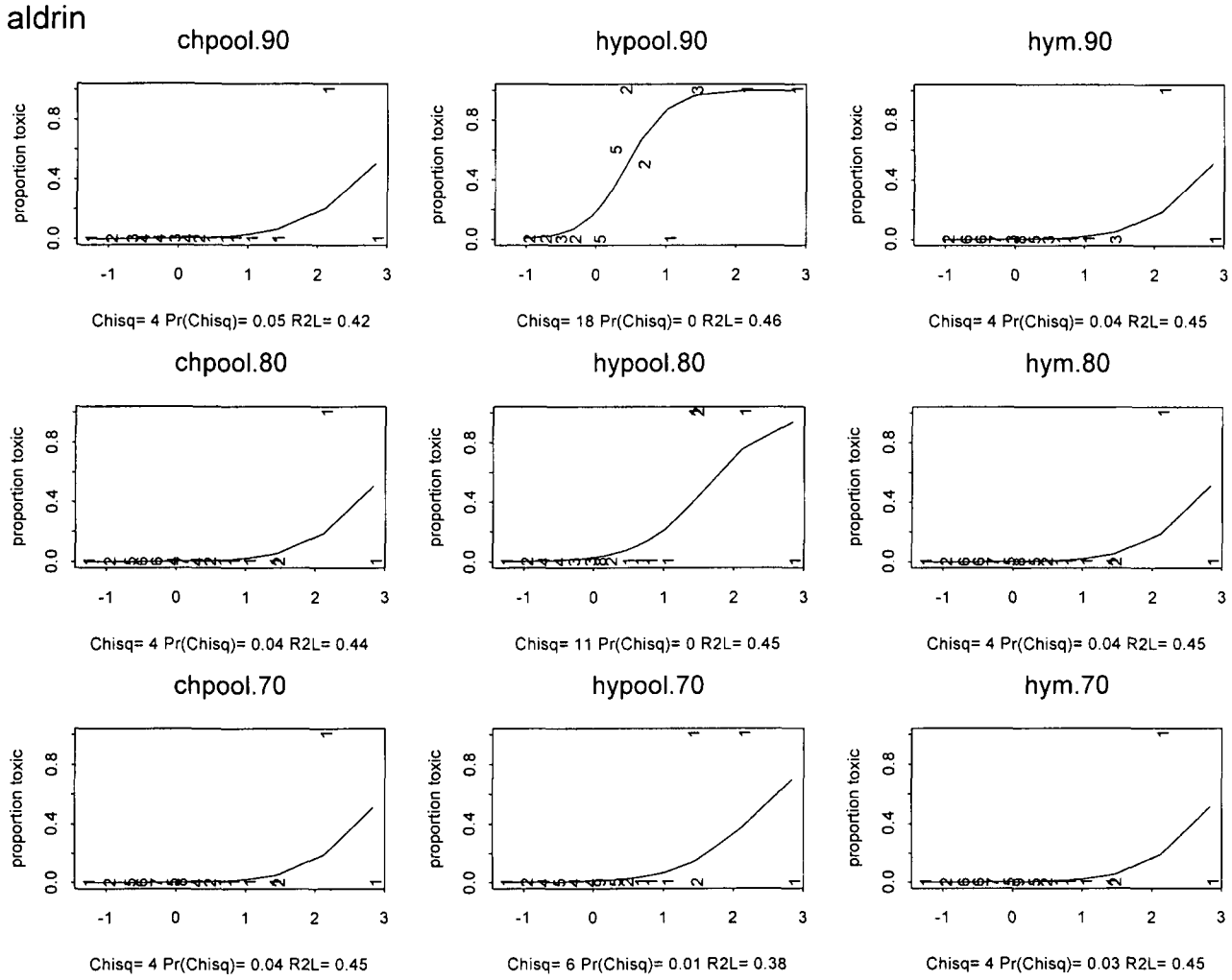
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-45. Logistic regression model – TEQ mammal (0.5 detection limit)



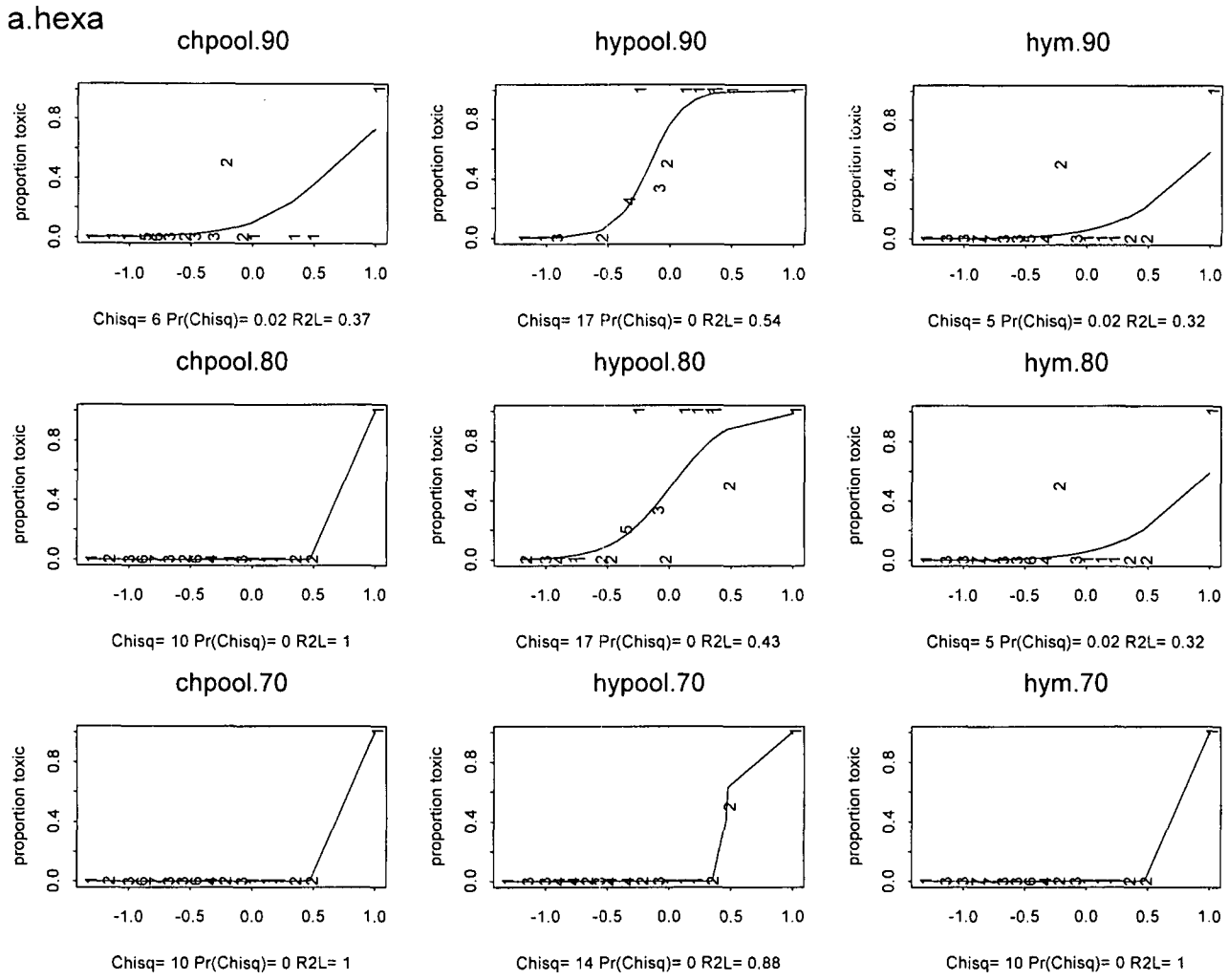
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-46. Logistic regression model – total PCBs



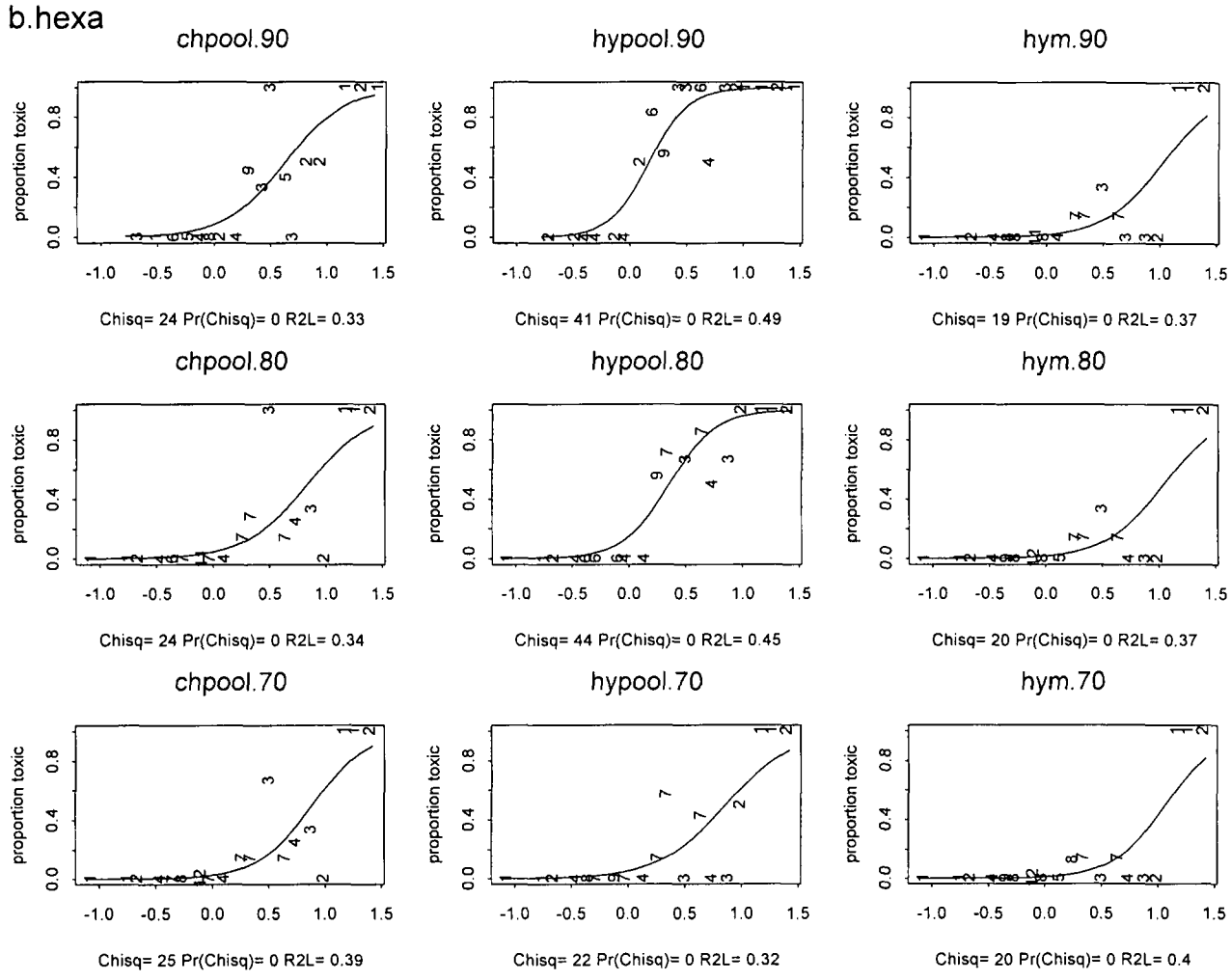
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-47. Logistic regression model – aldrin



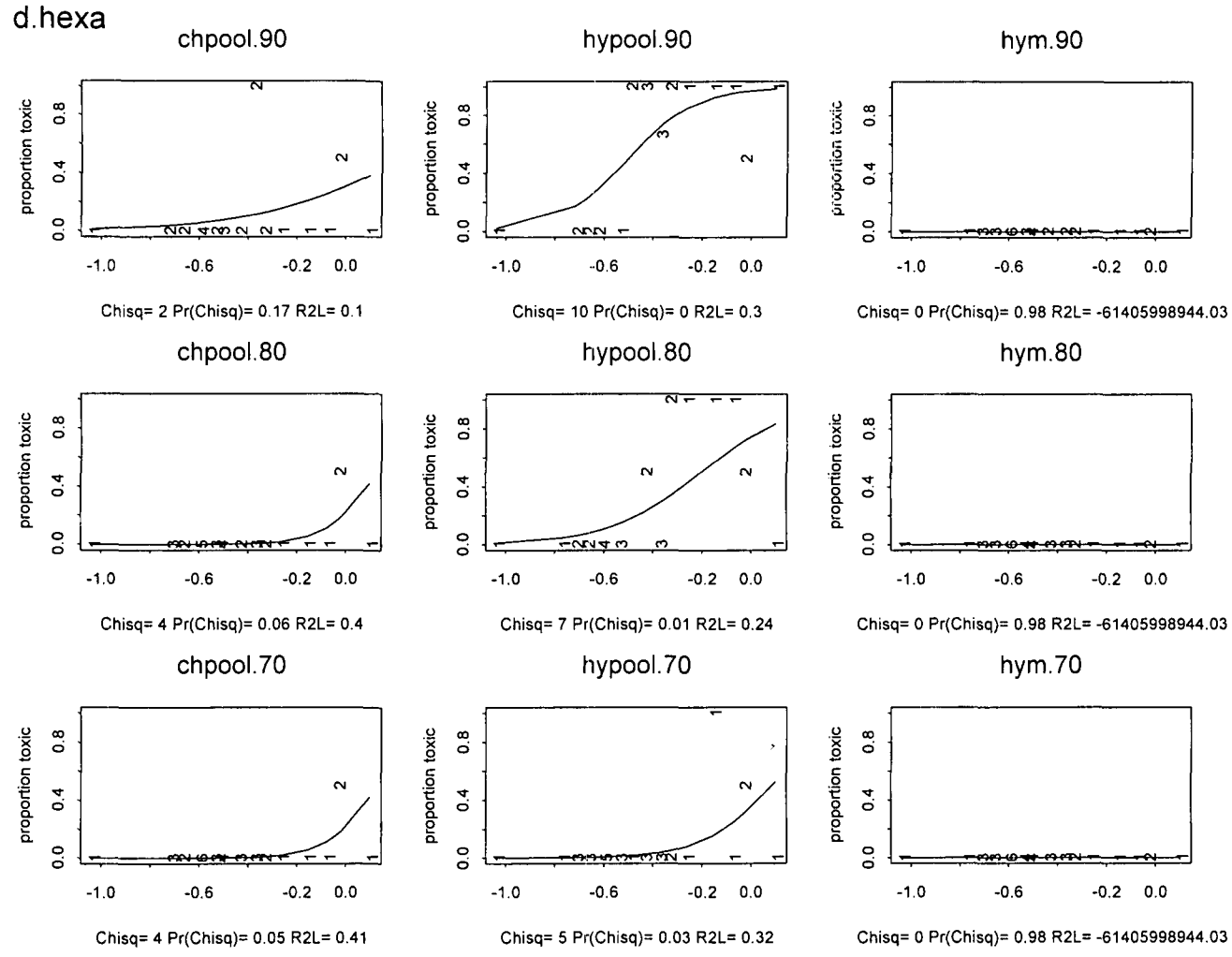
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-48. Logistic regression model – alpha-hexachlorocyclohexane



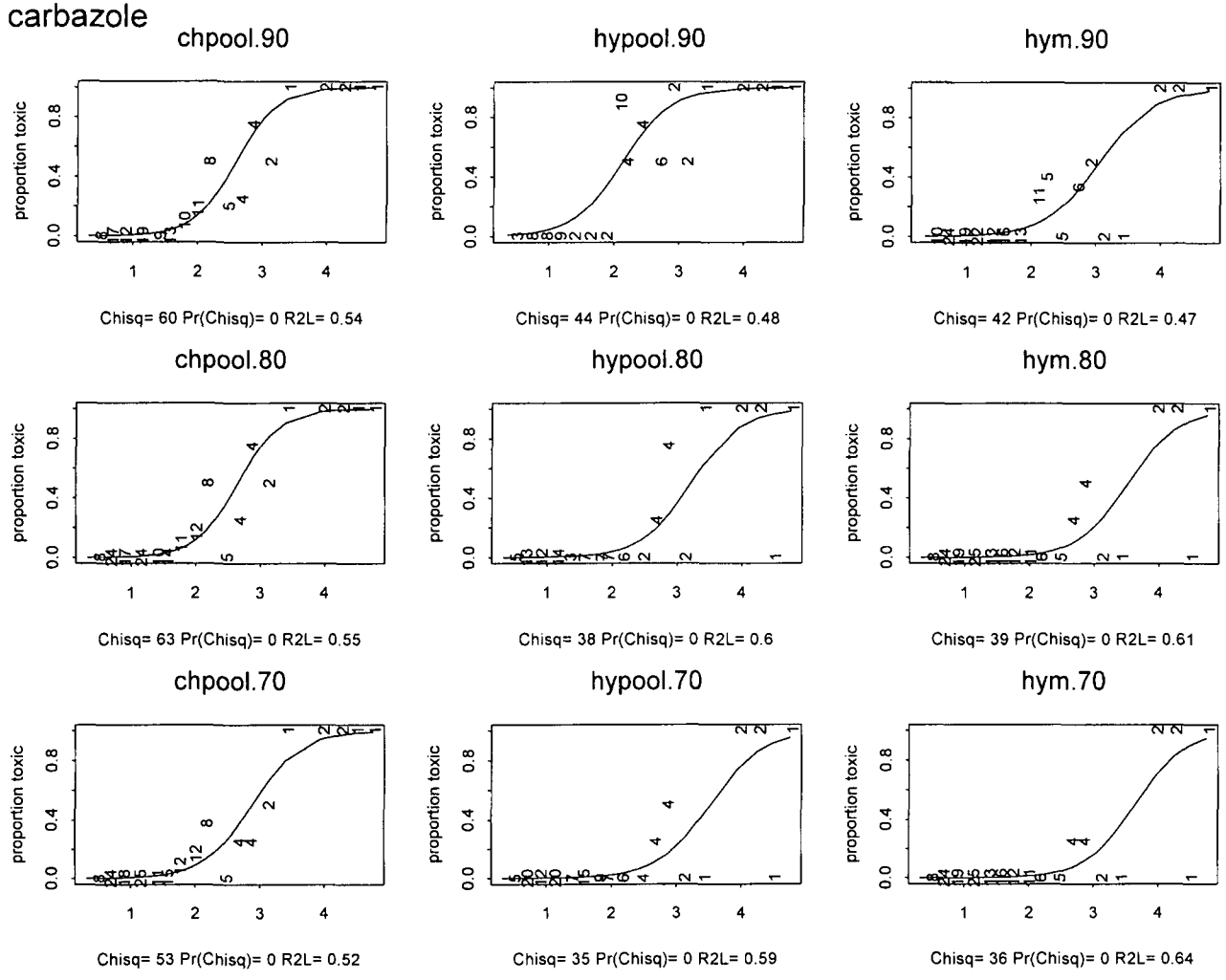
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-49. Logistic regression model – beta-hexachlorocyclohexane



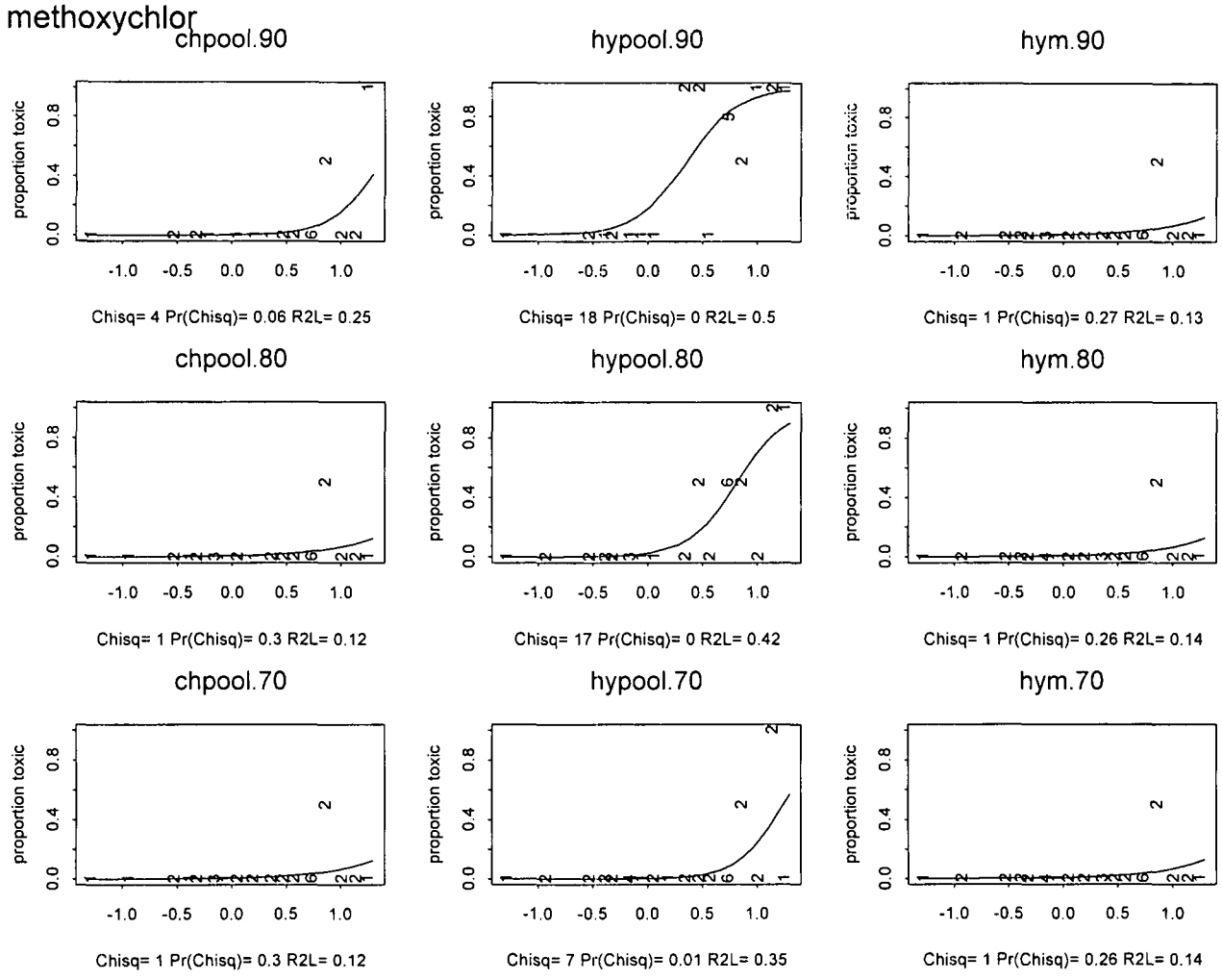
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-50. Logistic regression model – delta-hexachlorocyclohexane



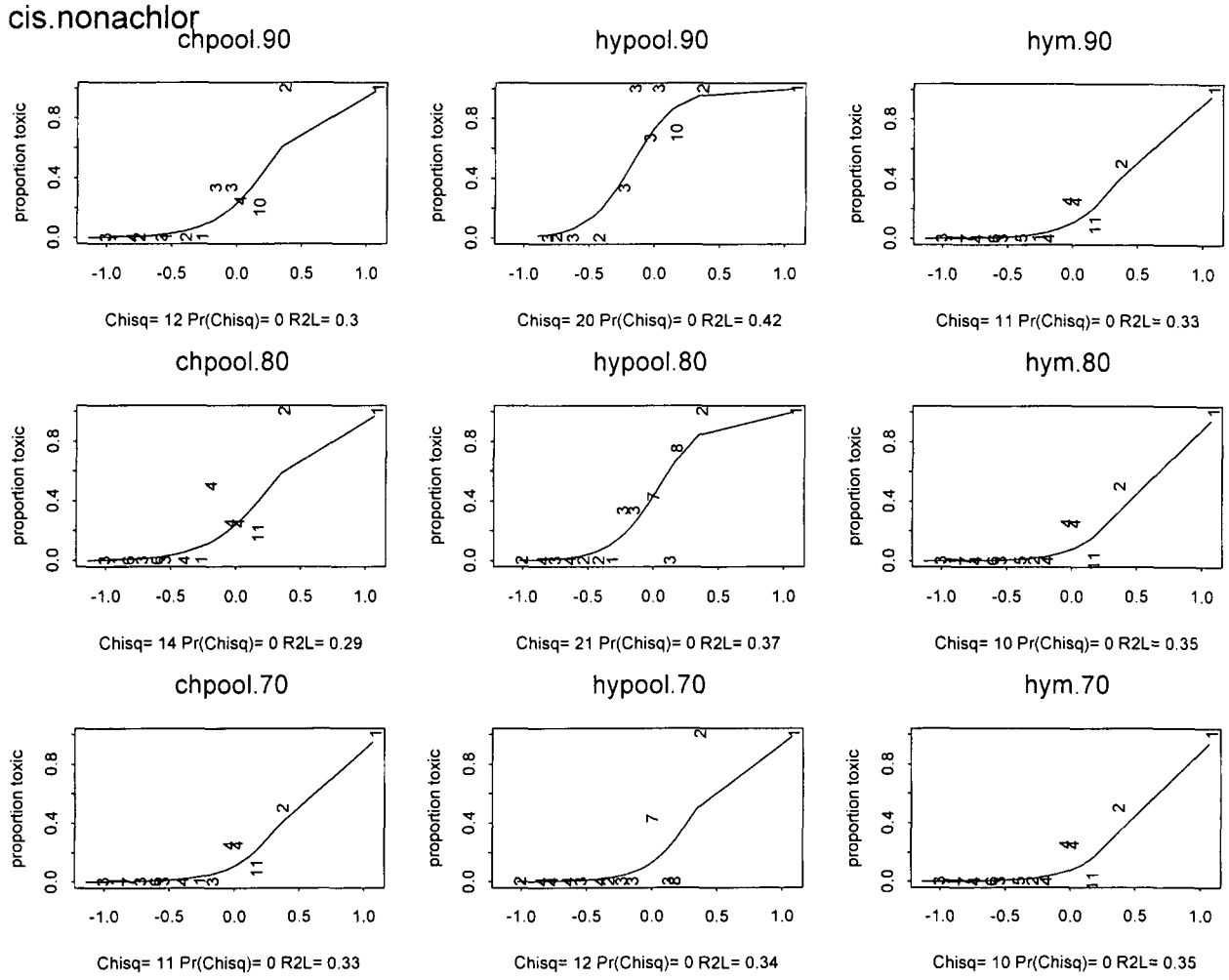
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-51. Logistic regression model – carbazole



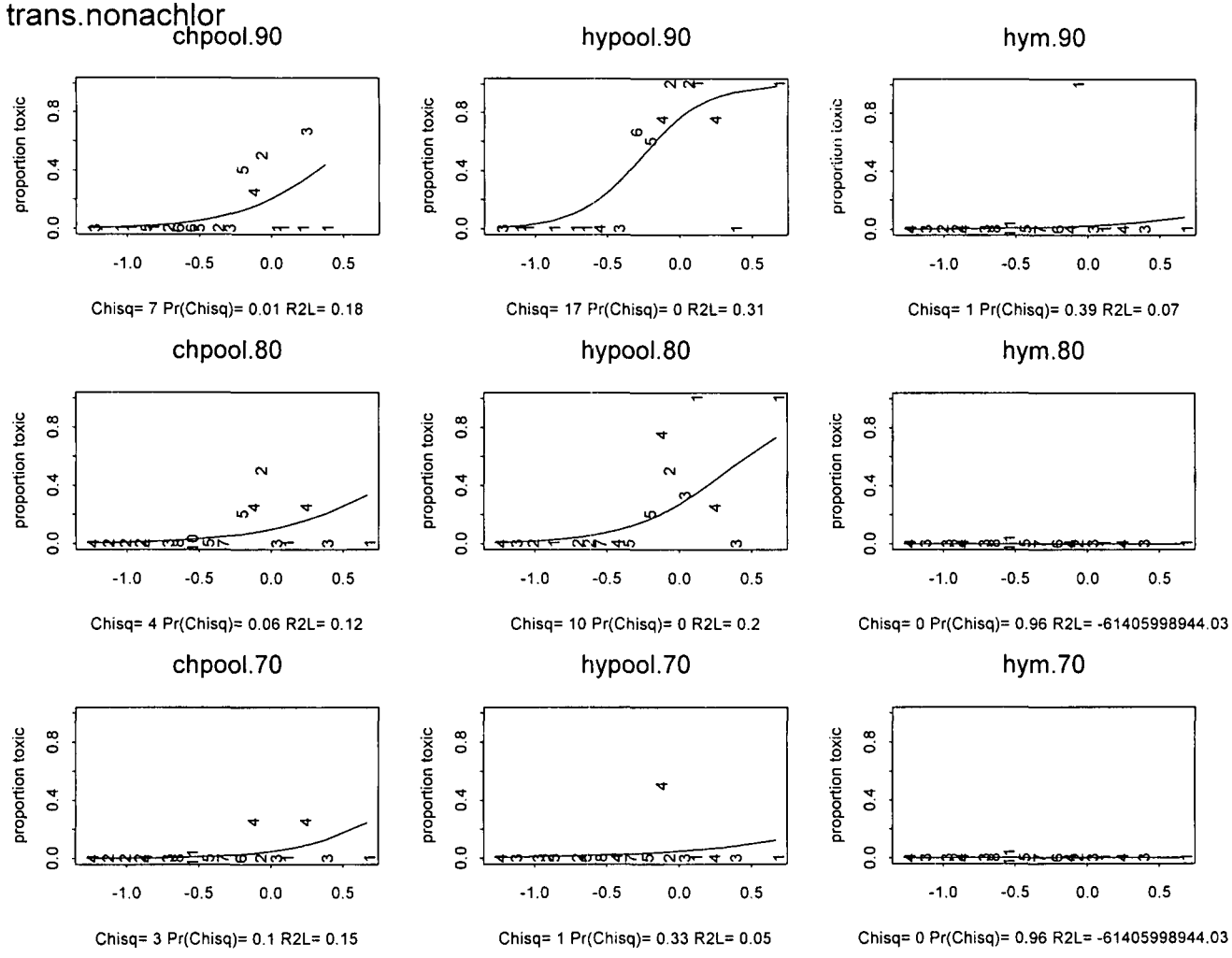
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-52. Logistic regression model – methoxychlor



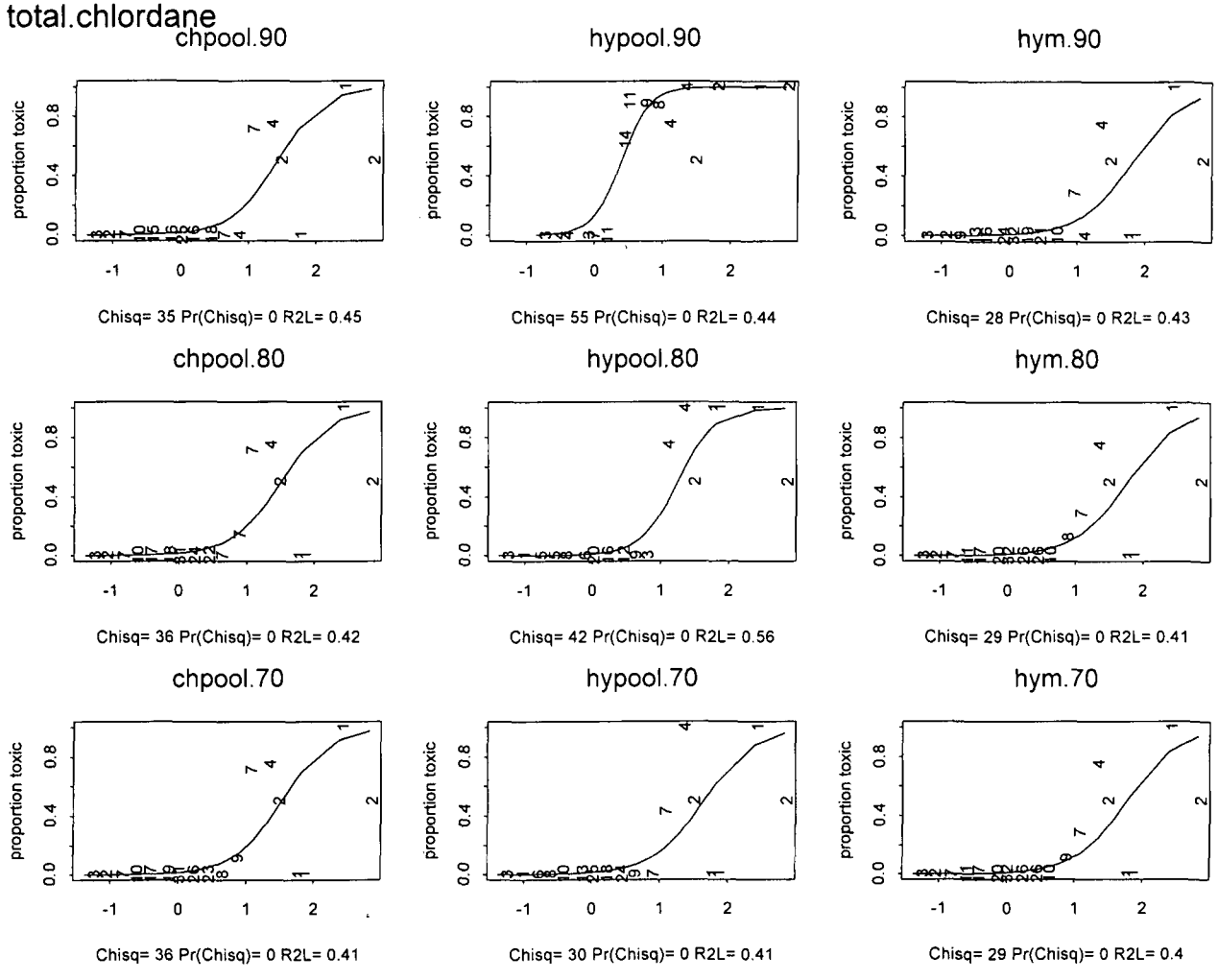
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-53. Logistic regression model – cis-nonachlor



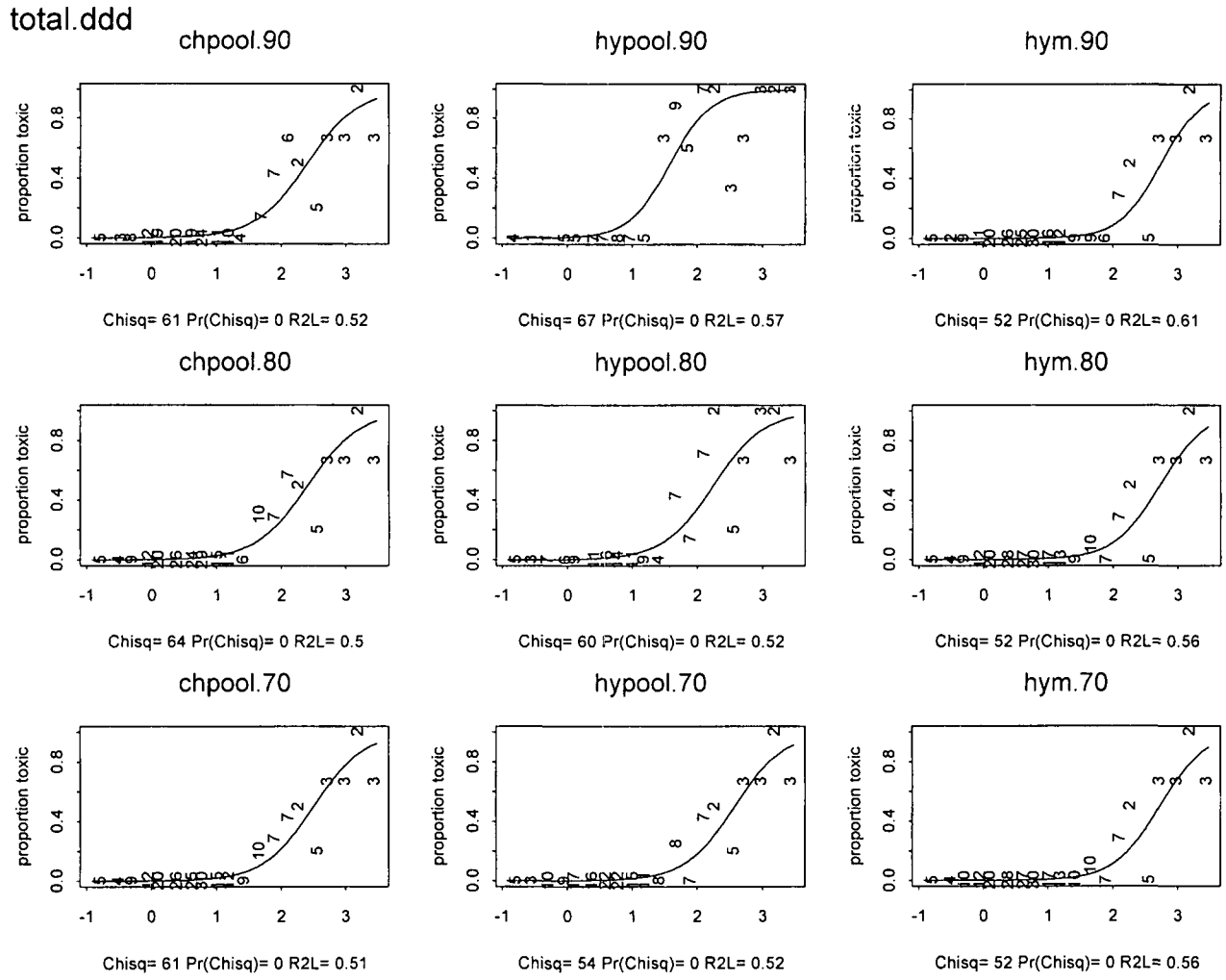
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-54. Logistic regression model – trans-nonachlor



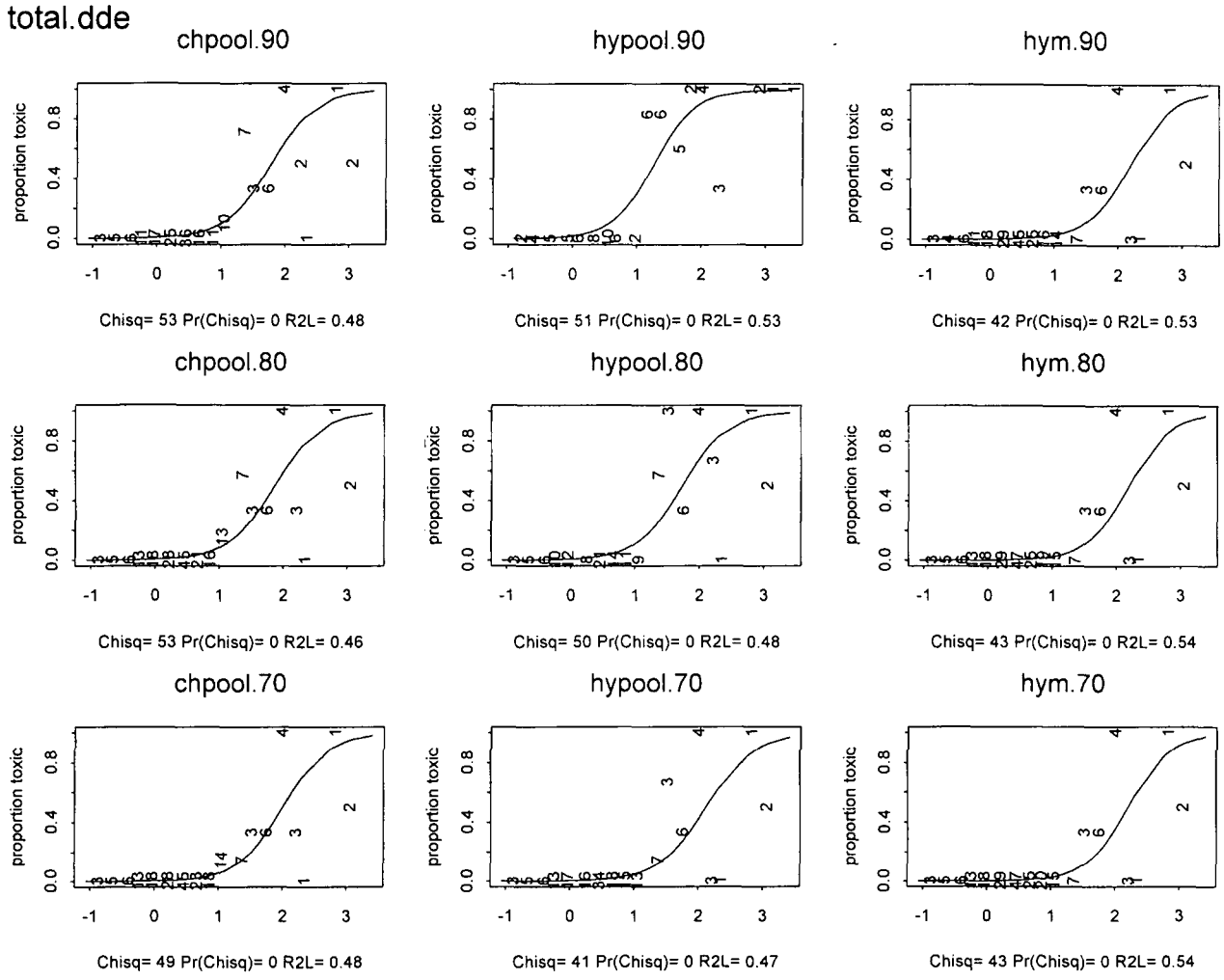
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-55. Logistic regression model – total chlordane



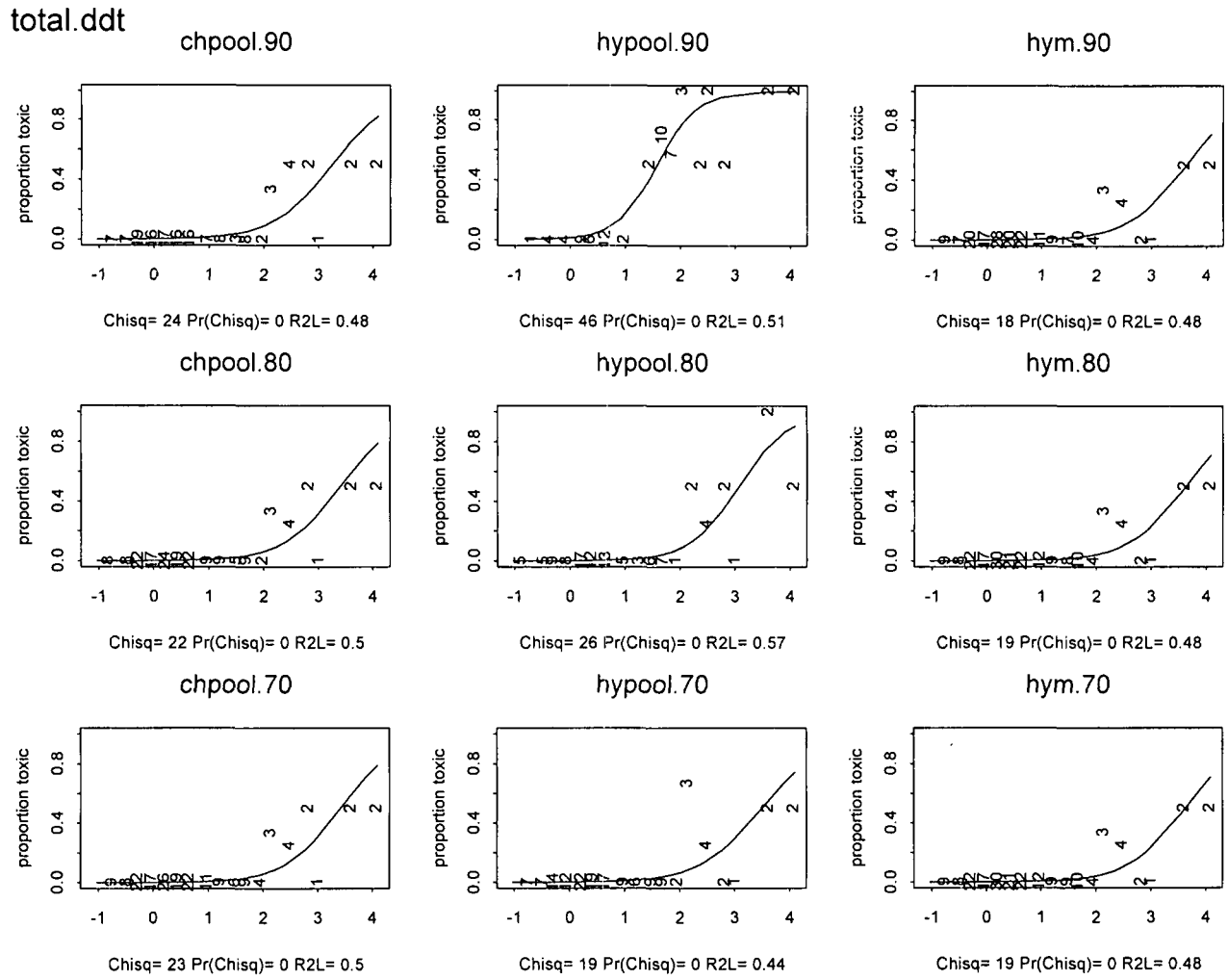
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-56. Logistic regression model – total DDD



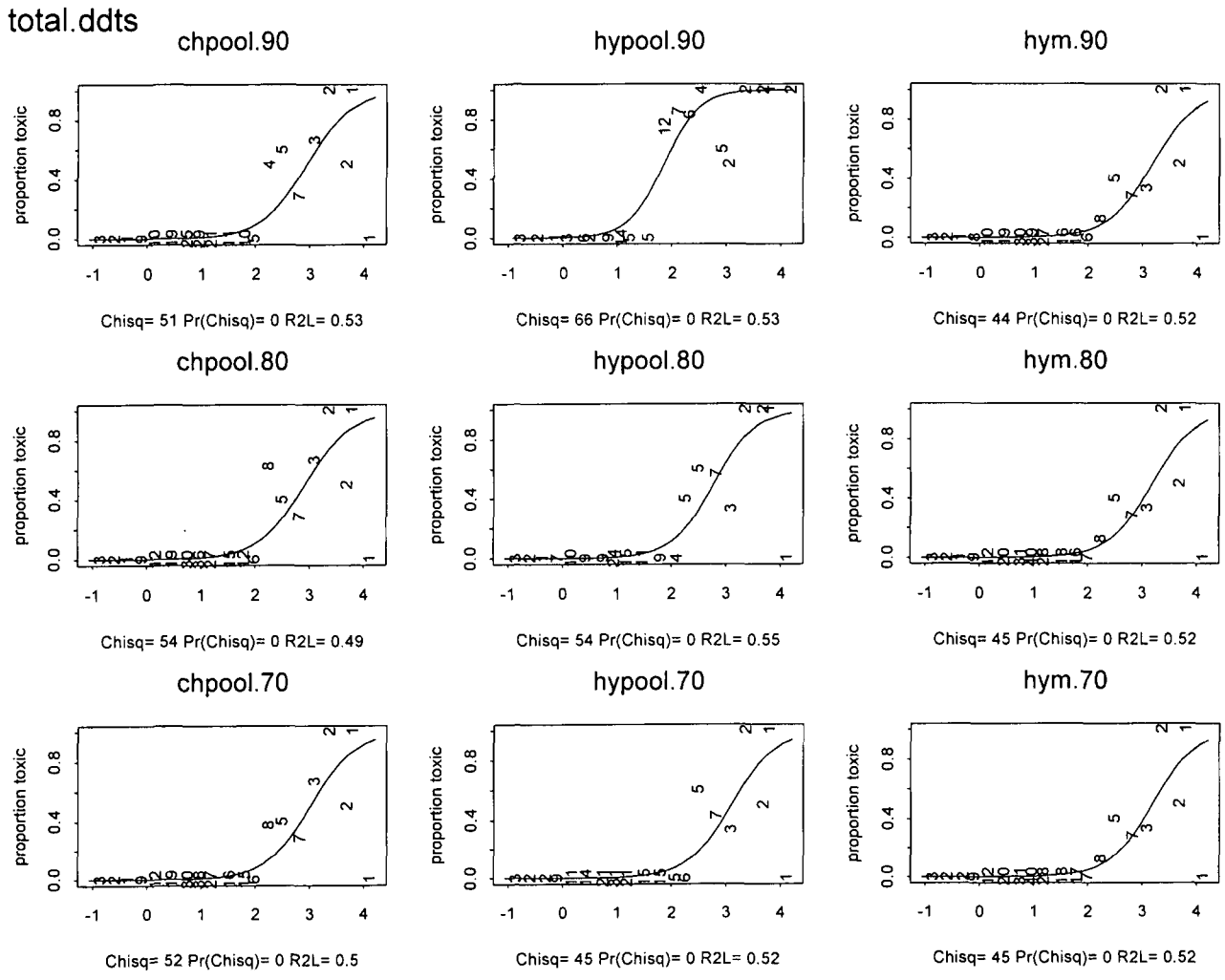
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-57. Logistic regression model – total DDE



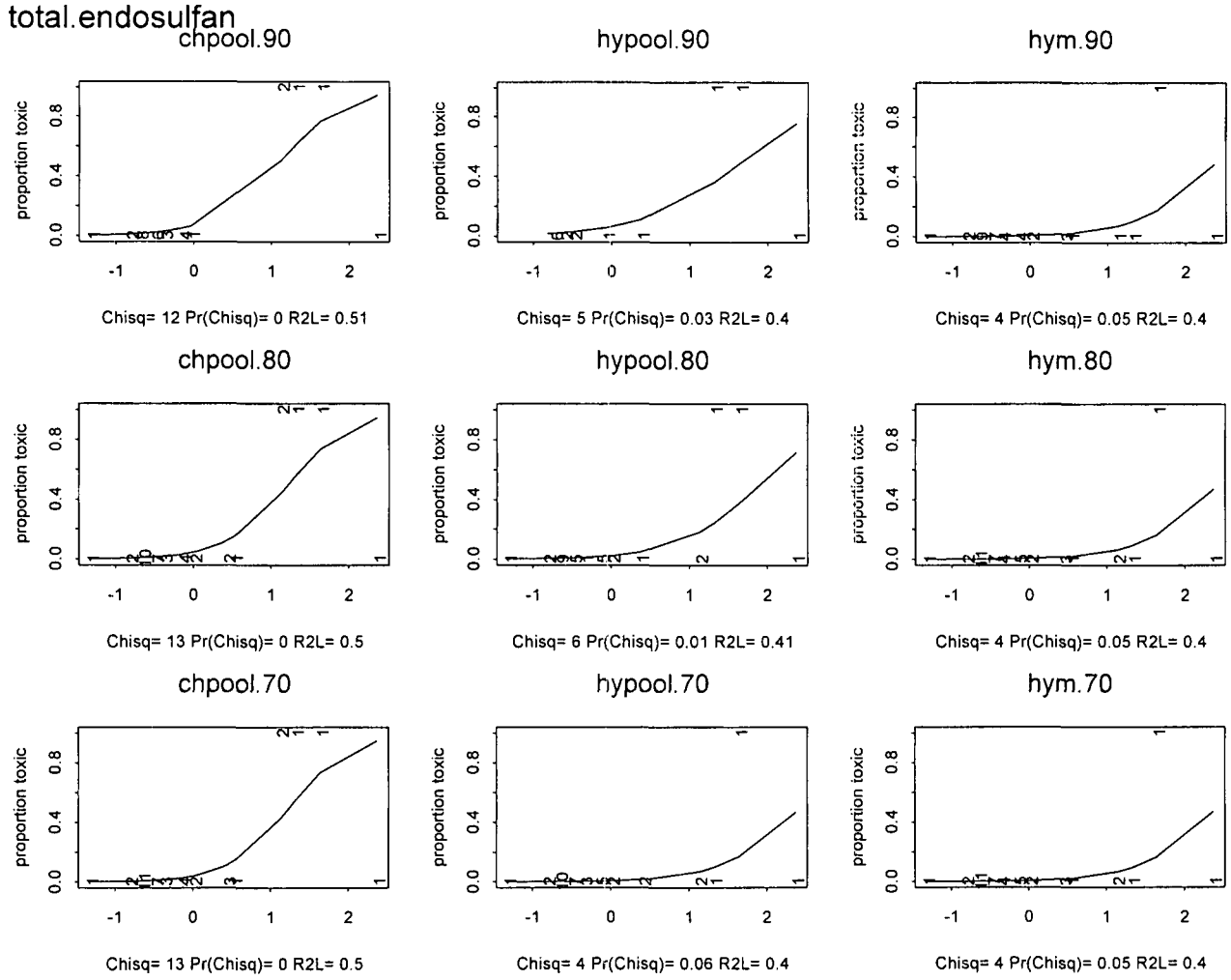
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-58. Logistic regression model – total DDT



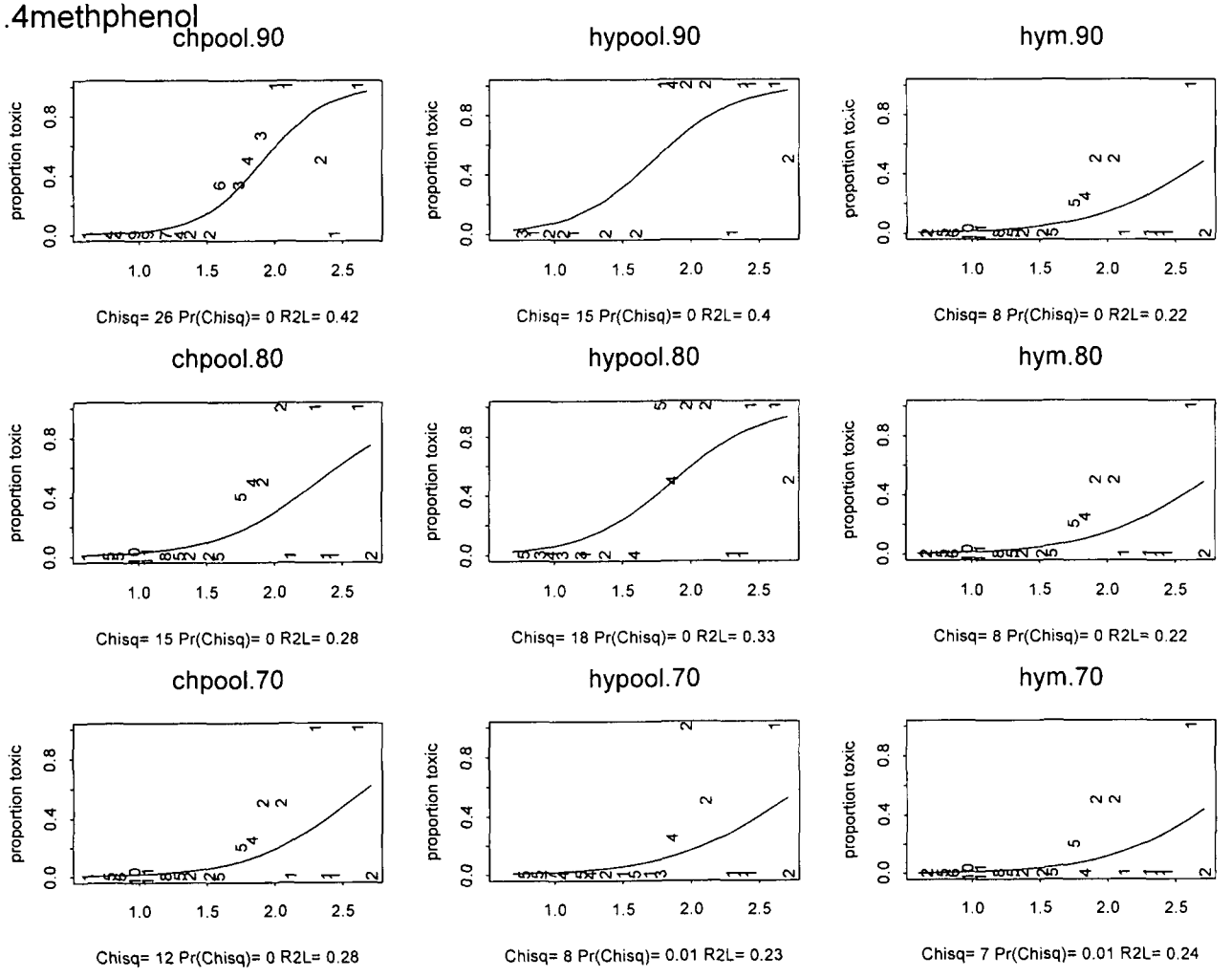
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-59. Logistic regression model – total DDTs



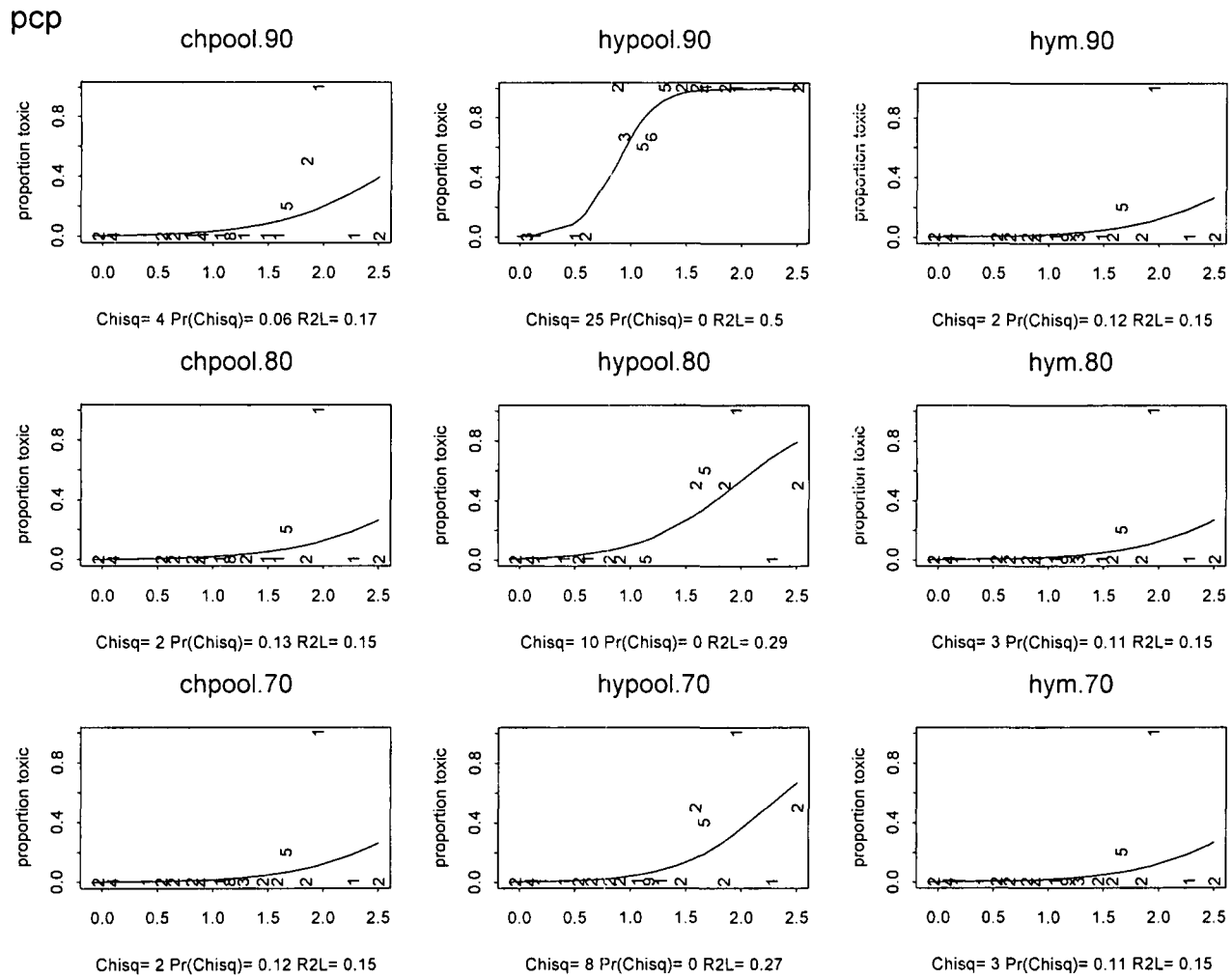
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-60. Logistic regression model – total endosulfan



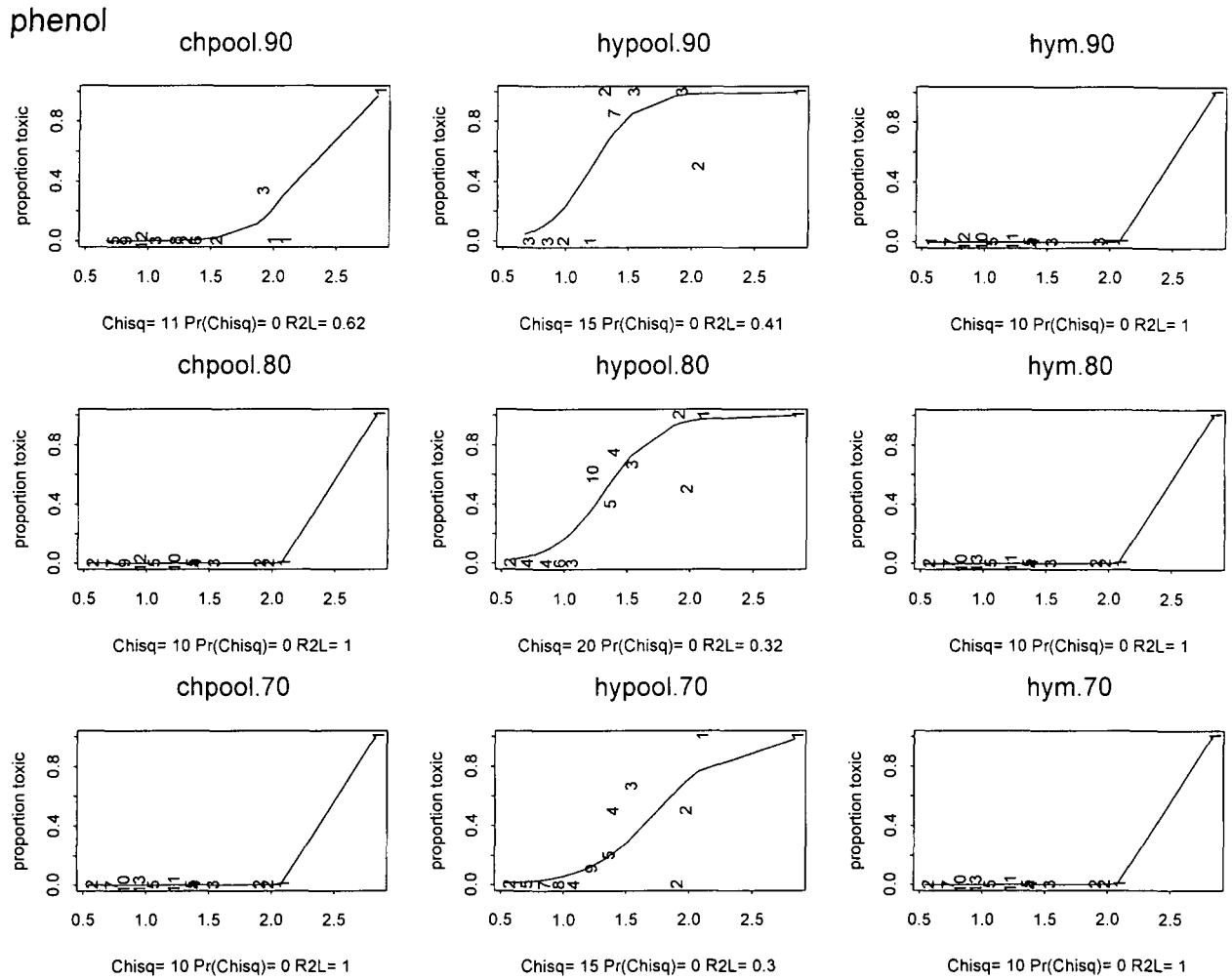
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-61. Logistic regression model – 4-methylphenol



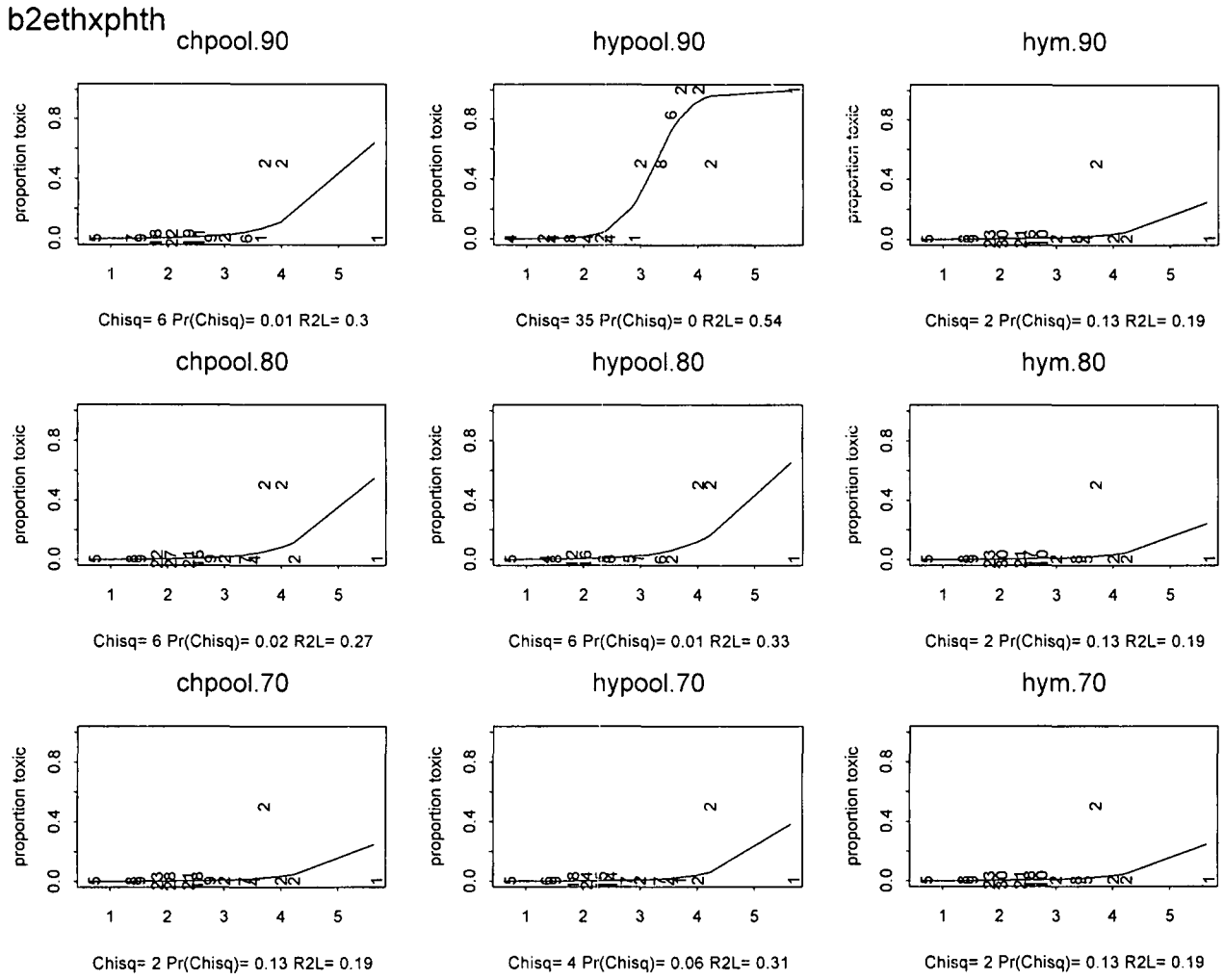
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-62. Logistic regression model – pentachlorophenol



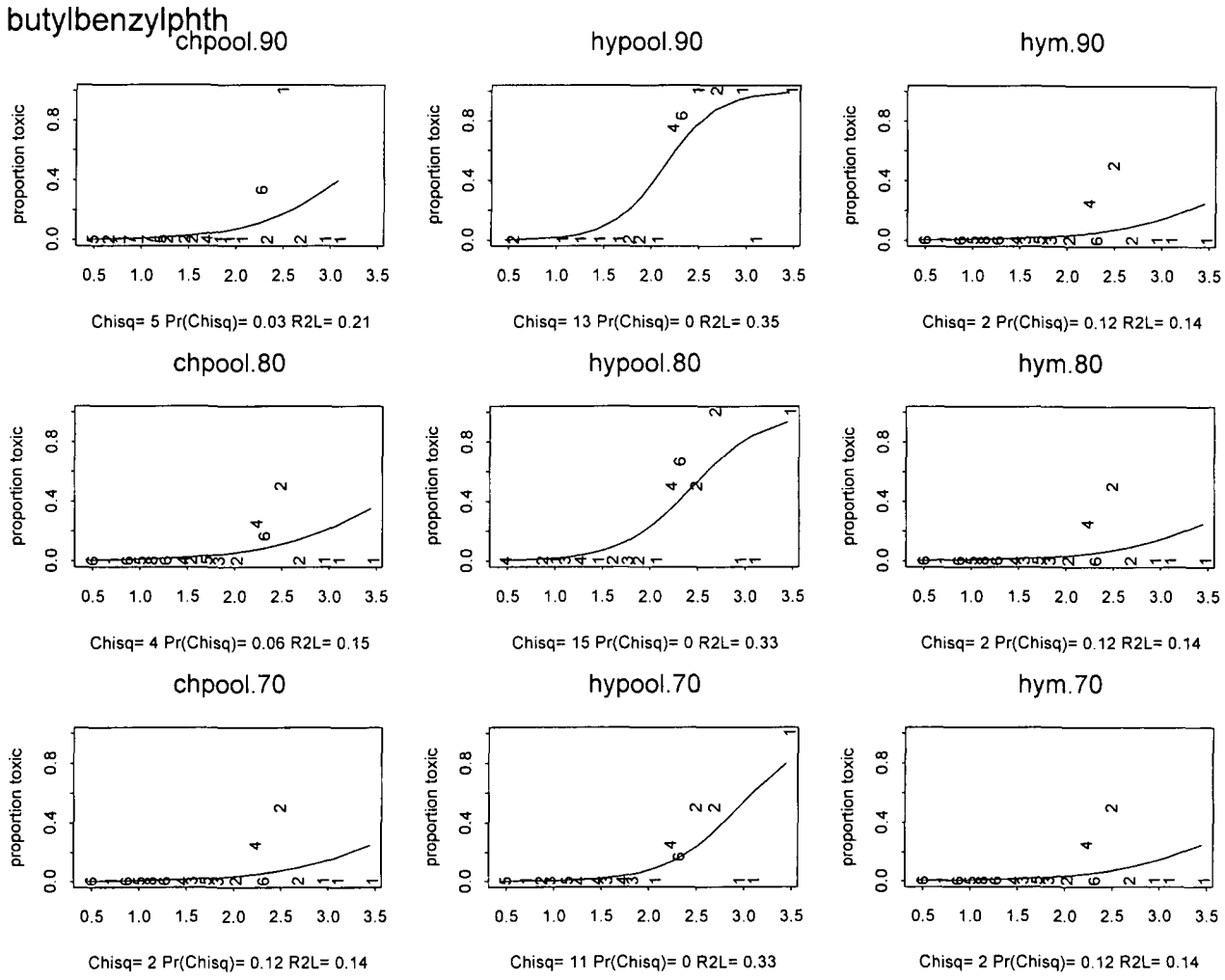
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-63. Logistic regression model – phenol



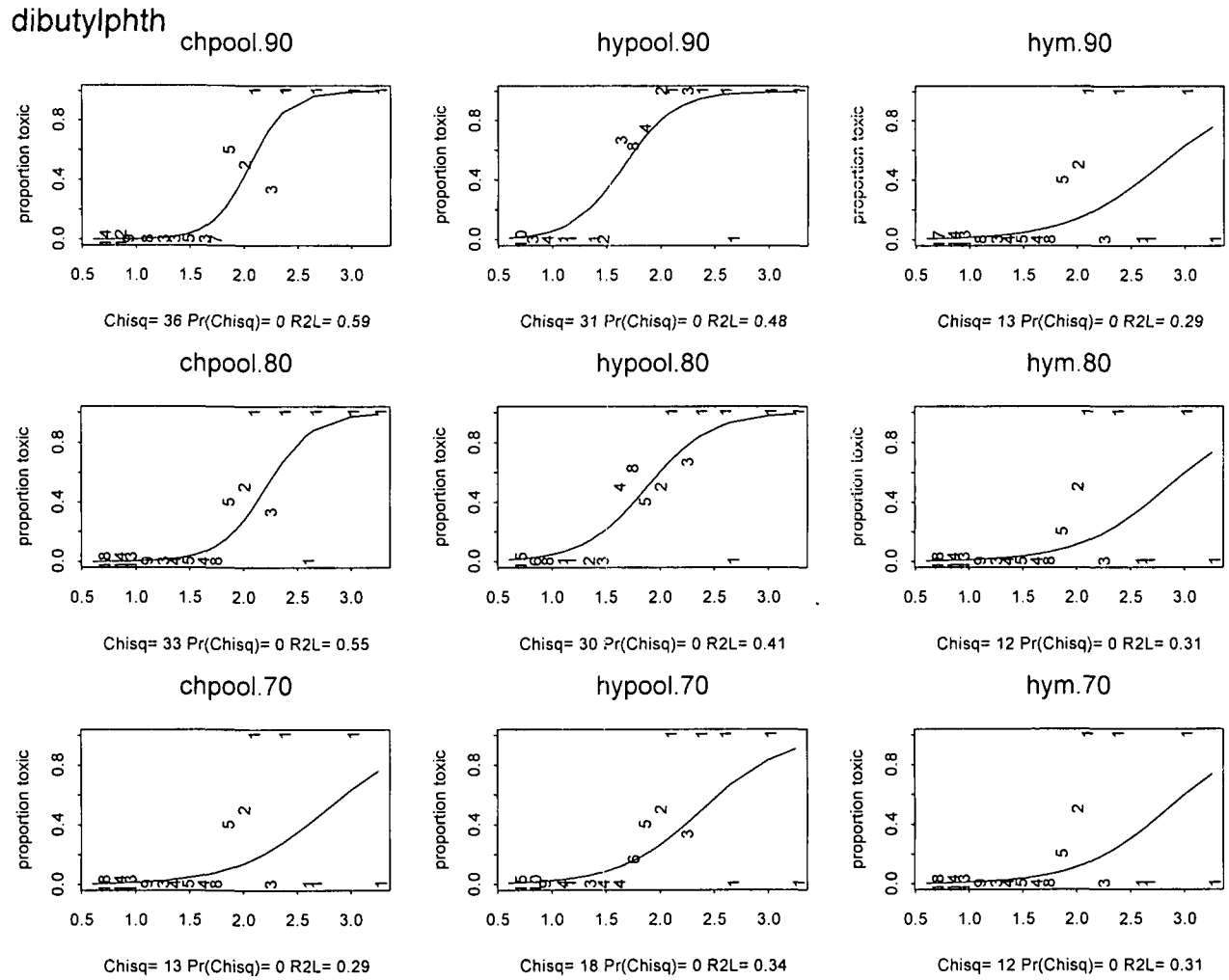
Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-64. Logistic regression model – bis(2-ethylhexyl)phthalate



Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-65. Logistic regression model – butylbenzylphthalate



Note: all plots in a column are for one toxicological endpoint (i.e., *Chironomus* pooled, *Hyalella* pooled, and *Hyalella* mortality), and all graphs in a row are for one effects level (L1 = .90; L2 = .80; L3 = .70).

Figure E-66. Logistic regression model – dibutylphthalate

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 10
1200 SIXTH AVENUE
SEATTLE, WA 98101

TARGET SHEET

The following document was not imaged.
This is due to the Original being:

_____	Oversized
X	CD Rom
_____	Computer Disk
_____	Video Tape
	Other: _____
	_____

**A copy of the document may be requested from the Superfund Records Center.

Document Information

Document ID #:	1309350
File #:	2.7.2-1
Site Name:	Portland Harbor (PORSF)

Portland Harbor Superfund Site Ecological Risk Assessment Interpretive Report: Estimating Risks to Benthic Organisms Using Predictive Models Based on Sediment Toxicity Tests Draft March 17, 2006